POLITECNICO DI TORINO
Repository ISTITUZIONALE

Participatory AI Justice in HCI: A Scoping Review

Original
Participatory AI Justice in HCI: A Scoping Review / Lupetti, M.L., Zaga, C., Cila, N.. - (2026), pp. 1-17. (2026 CHI Conference on Human Factors in Computing Systems, CHI 2026 Barcelona (ESP) April 13 - 17, 2026) [10.1145/3772318.3791704].

Availability:
This version is available at: 11583/3011817 since: 2026-06-09T12:37:56Z

Publisher:
Association for Computing Machinery

Published
DOI:10.1145/3772318.3791704

Terms of use:

This article is made available under terms and conditions as specified in the corresponding bibliographic description in the repository

Publisher copyright

(Article begins on next page)

20 November 2026

Participatory AI Justice in HCI: A Scoping Review

Maria Luce Lupetti
maria.lupetti@polito.it
Politecnico di Torino
Turin, Italy

Cristina Zaga
c.zaga@utwente.nl
University of Twente
Enschede, The Netherlands

Nazli Cila
N.Cila@tudelft.nl
Delft University of Technology
Delft, The Netherlands

Abstract

Participatory design is increasingly used to address the negative social impacts of artificial intelligence (AI), aiming for more inclusive and equitable innovation. However, it can inadvertently reproduce injustice and reinforce power imbalances, even with good intentions. While the HCI community is critical of these issues, it remains challenging for AI researchers and policy-makers to act upon these critiques. This paper presents a scoping review of Participatory AI research in HCI discussed through the lens of design justice. The goal is to provide a richer understanding of how current PAI work engages with justice and what the stakes and barriers are to putting justice principles in action. We conclude with raising methodological questions on the roles of researchers and partnership with communities, and the essential but instrumental role of artefacts in supporting knowledge production and social change. The work contributes to a holistic understanding of the current takes and stakes of Participatory AI in critical human-computer interaction research.

CCS Concepts

• **Human-centered computing** → **HCI theory, concepts and models**; **Accessibility theory, concepts and paradigms**.

Keywords

Participatory AI, Participatory Design, Social Justice, Design Justice, Artificial Intelligence

ACM Reference Format:
Maria Luce Lupetti, Cristina Zaga, and Nazli Cila. 2026. Participatory AI Justice in HCI: A Scoping Review. In *Proceedings of the 2026 CHI Conference on Human Factors in Computing Systems (CHI '26), April 13–17, 2026, Barcelona, Spain.* ACM, New York, NY, USA, 19 pages. https://doi.org/10.1145/3772318.3791704

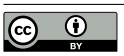
1 Introduction

> *"It is not coincidental that participatory design offers these challenges at a time when many businesses are discarding the conventional wisdom and struggling to remake themselves in the face of heightened international competition, and dramatic political change is taking place around the globe"* [92]

It was 1993 when Muller and Kuhn [92] mentioned this quote in their seminal article published by Communications of the ACM. They were witnessing times of significant change, both in corporate environments and the political landscape, that were opening up avenues for design to play a different role: that of a mediator, or even an advocate, for the needs of people for whom artefacts were being designed. Despite decades of participatory design work and a steady evolution of methods and theories around it, not much has changed. Today, more than ever, participatory design remains a key Human-Computer Interaction (HCI) approach to help navigate the muddy land that corporations and institutions shape, and that communities must walk through. We live once again in times of severe political instability, with wars taking place in areas from both the global south and the west [120], right-wing and conservative parties gaining momentum in more and more countries [15], our environment showing signs of collapse [95], and poverty rising globally [94]. Countering all this, there is an inebriating spring of "Artificial Intelligence (AI)"[1] development blooming globally. The number of large-scale AI models released has significantly increased since the debut of GPT-2 in 2019. By 2024, 167 large-scale models were launched [60], integrating AI into various applications. These advancements also link to such global challenges. As McQuillian [90] underscores, AI is an apparatus deeply entangled with our social, political, economic, and environmental infrastructures. To give a few examples, the widespread adoption of AI causes a surge in emissions and energy consumption [96], while also causing drought in areas that are already suffering from water scarcity [99]. AI-generated deep-fakes increasingly enter political elections and steer the public sentiment through misinformation [131]. In a similar vein, far-right parties often use AI-generated content to support populist narratives, e.g., anti-immigrant rhetoric [104]. AI-based decision support systems are showing a dramatic impact of algorithms on warfare, causing an increase in violence against civilians and an erosion of human accountability (see the current use of AI by the Israeli military force in Gaza [42]).

And last but not least, the adoption of AI in various public services and systems is manifesting issues of systematic inequality, biased treatment, and uneven distribution of resources. For example, the adoption of machine learning (ML) technologies in mortgage application processes was found to be carrying risks for black and Hispanic applicants, who are disproportionately less likely to gain credit [55]. Similarly, research has shown that the use of algorithmic diagnosis in chest radiographs can consistently and selectively underdiagnose underserved patient populations, especially those of intersectional underserved sub-populations, such as Hispanic female patients [112].

As these examples demonstrate, AI developments are closely linked to social, economic, environmental, and political issues that



This work is licensed under a Creative Commons Attribution 4.0 International License.
CHI'26, Barcelona, Spain
© 2026 Copyright held by the owner/author(s).
ACM ISBN 979-8-4007-2278-3/2026/04
https://doi.org/10.1145/3772318.3791704

[1] We use the term "AI" with quotation marks to denote its meaning in popular parlance, as suggested by the editorial guidelines of the *Critical AI journal*, but acknowledge the diversity of technologies that fall under the umbrella of the term AI [44]

need to be anticipated and mitigated. It has become crucial to consider who will be affected by AI implementations, and how, when, and where negative impacts may occur [125]. In this vein, *participatory design practices are gaining momentum again* under the idea that engaging the public, especially systematically excluded communities, will help challenge the cultural hegemony of AI [23] and ultimately lead to the development of algorithmic systems that positively impact society [5, 18, 20, 107, 140]. In fact, along with technical strategies to ensure more robust and accountable AI solutions, the critical AI community is calling for "Participatory AI (PAI)" to be integrated into AI development practices [18]) – a participatory turn designed to ensure that human values and rights become actionable design requirements for the development of AI systems [5, 10, 139] and that moral boundaries are respected [27]. This call also aligns with the recent views and strategies of numerous international institutions, such as the European Union's Special Interest Group on AI Ethics, as highlighted in the *Ethics Guidelines for Trustworthy AI* report. [68] argues for the participation of relevant stakeholders as a key practice to ensure diversity, nondiscrimination, fairness, and general acceptability of AI models and implementations. Similarly, the Organisation for Economic Cooperation and Development (OECD) argues that the implementation of AI in public sector institutions should come through an integrated approach that incorporates participatory mechanisms as a way to ensure alignment with public values and objectives [132].

However, as design scholarship warns us, participation practices are inherently subject to inclusivity issues and power struggles [10, 114]. Despite its long tradition and growth in methods and scopes, participatory design consistently fails to address how it perpetuates historically imbalanced power dynamics [10, 16, 19], and how it leaves unattended the radical issues that marginalized communities experience daily (i.e., complexities associated with the location they live in, language and knowledge barriers, financial burdens, time and more [46]), hindering their inclusion in design processes [10, 64, 114]. The growing popularity of participatory practices in AI innovation contexts is likely to exacerbate these issues. The participatory ambition to be inclusive and open to public deliberation clashes with the configuration and dynamics of the current AI landscape, which is characterised by *centrally-owned-yet-globalised* infrastructures governed by tech giants with explicit capitalist ambitions [143]. The focus on localised knowledge from community engagement contrasts with the goal of foundation models to be versatile across different domains and regions [124]. Moreover, the resources and time needed to engage with communities clash with the pace of technological innovation, which cannot afford the practical implications of community participation [20, 36]. As a result, the majority of PAI efforts end up being performative [33, 114]. Stakeholders are largely involved to provide input on discrete implementation parameters, rather than making key AI design decisions [36]. Experts do not even take the needs or recommendations of the different stakeholder groups into full consideration [114]. Participatory engagements are primarily utilised to improve AI systems and are often exploited by corporate entities to maximise profits [18].

Critical HCI and AI scholarship is increasingly aware of these issues. Scholars have been calling for *justice-oriented* approaches to participatory design, namely approaches that foreground structural inequality, epistemic inclusion, and redistributive outcomes. There are multiple studies aiming to involve communities throughout the whole AI development lifecycle [101, 142] and critical reviews aiming to build literacy on how matters of justice are entangled with participation processes [33, 36, 50]. Yet, enormous variations remain in terms of methods and approaches used to achieve participation [36], which too often result in being performative rather than truly empowering for communities. It remains challenging, especially for AI practitioners –developers and regulating bodies alike– to find and use the body of knowledge that the HCI community, and the critical AI subcommunity in particular, produces. We take a similar stance: having to initiate applied research on PAI and social justice, we find ourselves negotiating between the ambition to embrace our social justice ideals and values, and the practical constraints of performing participation between the constraints of funded research, with a limited time span and resources. Therefore, in this work, we aim to build a deeper understanding of where the HCI/critical AI community is at, when it comes to practising justice-oriented participatory AI.

We present a scoping literature review that examines PAI articles published in leading HCI conferences and journals. This review highlights and reflects on contributions that adopt critical perspectives on PAI work and its social justice implications. In particular, we adopt a relational and intersectional understanding of social justice, grounded in the Design Justice Principles [40], which foregrounds the lived experience and leadership of communities most impacted by socio-technical systems. Following Fraser [54], we understand justice as not only about equitable distribution, but also about recognition and participatory parity, ensuring all people have an equal stake in shaping the systems that affect their lives. This framing informs our use of the concept *Participatory AI Justice*, which we take up as a way to interrogate whether participatory methods in AI reinforce or resist structural inequalities.

The goal of this work is to provide a richer understanding of how current PAI work engages with justice, whether it offers practical resources to put justice principles in action, and what the challenges are to operating in this space. We establish a foundation for future justice-oriented PAI practices that illustrates how participation in AI can occur –and what is at stake. We contribute with 1) a holistic understanding of participatory AI practices through the lens of design justice; 2) a set of methodological reflections for practising justice-oriented PAI work; and 3) a discussion of future research opportunities.

2 Related works

The AI community is increasingly embracing participatory practices to mitigate adverse impacts of AI [145] and help develop systems that can be beneficial for society [87]. By incorporating the needs, perspectives, and concerns of the broader public, including communities that are systematically excluded, into the design process [18, 20], PAI practices increase the social responsibility of AI and potentially reduce bias both in the logic and data through a collaborative decision-making process [48]. As Birhane and colleagues [18] argue, participatory approaches in AI should not merely capture individual opinions, but instead focus on fostering inclusion,

embracing diverse perspectives, ensuring collective safety, and promoting shared ownership. This perspective reframes the dynamic between designers and users, encouraging a shift toward more equitable collaborations where all participants act as co-designers and co-creators.PAI plays out in different formats and moments of the AI development lifecycle. These range from public debates with laypeople to anticipate risks [93], to technical approaches to support public expression of preferences about AI systems for the public sector [141], as well as approaches in between [61]. The AI community (including engineers, designers, industry professionals, HCI researchers and more) largely considers PAI as an approach that should continue to grow and be refined as a key component of the AI development and deployment lifecycle as a way for communities, especially those at the margins of society, to contribute to strategic decisions around AI [18]. In this perspective, PAI centres ambitions of social justice in place of capitalist ideals of progress and interests of profit. However, PAI is also a methodological space with inherent complexities and contradictions [87]. In what follows, we unpack the risks and complexities of participation, both in the participatory design tradition and in the current critical AI literature.

2.1 Participatory Design and its social justice issues

Participatory Design (PD) is an approach that "*questions major assumptions about technologies in workplaces, communities, homes, and social institutions*" [92]. Drawing on interdisciplinary traditions of science and technology studies, sociology, psychology, and design, PD integrates a diverse range of methods and tools (e.g. worksheets, prototypes, product representations, event schedules, mock-ups, and other design things) to collaboratively envision preferable futures [84]. It opens the innovation process to be a joint effort between people from different organisations and backgrounds, up to the point of sharing and combining ideas – commonly referred to as *co-design* [121].

PD originated within the Scandinavian workplace democracy movements and was shaped by critical approaches to the work-related technologies, particularly computer-based systems. These contexts highlighted issues of power distribution in the workplace over time [78]. The approach was initially driven by practical motivations–a need for gathering experiential knowledge about artefacts in use, which led to a breadth of works looking at "*the tacit, invisible aspects of human activity*" [117], with the ultimate intent of developing products that better align with user needs and skills [70]. Due to its practical value in facilitating innovation processes, PD has gained popularity and is now used in various fields, including entertainment technologies, urban planning, assistive robotics, and automated vehicles. In its highest ambitions, however, PD is not only practised for its utilitarian outcomes, but also for its potential to stir socio-political change [92]. Underlying PD, in fact, there is an ambition to dismantle power imbalances between "*those charged with technology design and those who must live with its consequences*" [78]. As such, PD is also seen as a practice that is "*antithetical to consumerism*" [108], and that opens up productive dialogues and cooperation among diverse stakeholders carrying contrasting values. As PD then shifts the focus from individual projects to the complex net of relations that exist around the development of products, it expands our understanding of design to include its role as facilitator of organisational, legal, social, and political change [78]. The operational space of PD expands along its nomenclatures (i.e., Community-based Participatory Design [39]), and finds increasing adoption in justice-oriented movements like the *Design Justice Network*[41], especially as it harnesses collective power and experiences as a way to achieve safer, more just, more accessible, and more sustainable worlds. From initial considerations of when to employ these practices, who to engage, and how many at the same time [92], PD has now become a space where researchers and practitioners question their own identities, to understand how their actions relate to issues of equity and justice. Designers reflect on their privileges and positions of power [64, 144], and question their role in light of the growing awareness that **design is a universal human activity** [34]. PD encourages a radical rethinking–even a downsizing– of the role of the designer, who is better seen as a facilitator rather than a decision-maker [40, 64, 108], and a shift in innovation practices. PD questions the centrality of technological progress up to the point of making spaces for resistance [4] and refusal [102].

Despite these noble intentions, however, PD often fails to enact meaningful participation and continues to grapple with the very justice issues it aims to address. As Lodato and DiSalvo [84] argue, especially when applied in the public realm, PD is performed within invisible but strict constraints. These include differing ideals and values of stakeholders, inadequate administrative and legal frameworks to support the envisioned possibilities, and practical limitations on the range of possible experiments. At the time of participatory engagements, *design decisions are often already taken*: the *actual* design happens elsewhere at other times, i.e., when a technology is developed. The PD processes end up being actions for smoothing the deployment of AI systems [84]. Understanding these constraints is crucial as they shape design processes, highlighting how design can perpetuate inequalities and power structures [30]. Inequalities, in fact, manifest at all levels of the design process [34] and are particularly exacerbated by a specific design culture: one that stems from a privileged, white, youthful, and upper-to-middle-class approach to innovation [64]. In this regard, Harrington and colleagues [64] highlight that design workshops are not neutral spaces; they are shaped by specific social and cultural norms that can inadvertently exclude or marginalise certain participants. Participation often lacks a thorough understanding of the historical context that communities live in and are generated from. As a consequence, it can devalue and reduce the lived experiences of those who may not fit the 'norms' of participants in a dominant design culture, and can draw further distinctions between participants and researchers [64]. This adds up to the already acknowledged complexity of involving people in ways that are most conducive to their ability to participate [108], and the need for finding ways to give voice to everyone's lived experience as unique and brilliant contributions to design [40].

Therefore, it remains an open challenge for designers to practice PD equitably [115], especially when engaging underserved communities [64]. Designers must interrogate themselves about their positions of power and adopt practices that not only support and uplift communities but also actively resist systems of exploitation

and oppression [40]. A crucial starting point is accounting for the history and context of where participation happens. Doing so is essential for building trustworthy processes that challenge corporate design thinking traditions and making space for inclusive processes that honour the complexity of people's experiences, views, and aspirations [64].

2.2 Current critiques to Participatory AI

As seen in other sectors where PD has been implemented, PAI efforts carry the risk of superficial engagement, in which community involvement is **tokenistic** rather than substantive [103]. Although often rooted in well-meaning intentions, such efforts can result in "participation washing" instead of fostering meaningful engagement [47, 103, 114]. The underlying causes are complex. First, there is a tension between the desire for inclusivity in public deliberation and the current AI landscape, which is dominated by centralised infrastructures owned by tech giants with capitalist goals [143]. Secondly, there is a tension between the PD's commitment to localised and context-specific knowledge gained from engaging with communities and the aspiration for foundation models to be versatile and applicable across various domains and geographies [124]. These tensions are compounded by two persistent barriers: the significant resource and time demands required to meaningfully implement PAI, which often clash with the rapid pace of tech development [20, 36], and the inherently extractive nature of AI itself.

As researchers grapple with the tensions between ideals of inclusion and the realities of AI's extractive infrastructure, several have begun to conceptualise different forms, levels, and meanings of participation in AI development. The following sub-sections explore these evolving understandings by mapping how participation is being theorised in the current PAI landscape, including its conceptual, institutional, and practical limitations.

2.2.1 PAI challenges prevailing understanding of participation in design. The problems mentioned above have recently motivated researchers, from academia and beyond, to build a deeper understanding of participation within the space of AI. The Ada Lovelace Institute, for instance, has defined a five-tier participatory mechanism that ranges from inform, consult, involve, collaborate, to empower (as cited by [36]). In a similar vein, Corbett and colleagues [33] unpacked potential PAI dynamics by revisiting Sherry Arnstein's *Ladder of Citizen Participation* [7]. This is an eight-rung ladder, where each rung depicts the degree of power afforded to people, spanning from nonparticipation, through tokenism, to citizens' power. Sloane and colleagues [114] offer a complementary framework that introduces justice as a central dimension in PAI. They distinguish between two main participatory modes: *participation as consultation*, where engagement is typically short-term and limited to specific design questions (e.g. through workshops or dialogic sessions) and *participation as justice*, which emphasises long-term, collaborative partnerships rooted in social justice traditions. These include approaches informed by crip technoscience, intersectional feminism, and decolonial theories, foregrounding "*designing with*" marginalised communities throughout the whole development lifecycle as a way to ensure that outcomes hold social and political value. Importantly, the authors also underscore that AI systems are inherently participatory, though not in ways that align with democratic ideals [113]". AI relies on the continuous labour of ordinary users, such as annotators, auditors, and testers, whose data and feedback fuel model performance assessment [3]. Furthermore, as models continue to scale by getting trained on ever larger swaths of human-generated data and "fine-tuned" through processes involving human feedback in some form, participation ends up being controversially entangled with optimisation practices. Acknowledging the crucial role of annotators in the AI ecosystem, critical AI scholars have started to explore how can annotators can be regarded as bearers of situated knowledge and valued for their potential role to achieve contextually relevant value alignment of large language models [8, 79]. Despite these (few) promising propositions, however, it remains unclear how to enable meaningful and intentional contribution to AI development when participation takes the form of work or hidden data extraction mechanisms.

These forms of engagement may be even better understood as '*non-participation*' [33] as people in these practices do not have a real say in the decision-making around AI development [36]. But even when participation is well-intended and people are involved with the ambitions to shape better systems, the overarching AI infrastructure limits their capacity to contribute meaningfully. Especially in the case of foundation models, there is an inherent limit–what Suresh and colleagues [124] call "*participatory ceiling*"–on the ability that impacted communities have to meaningfully shape an AI system, as this is intended to be almost universally applicable.

2.2.2 PAI builds on a tradition of institutional and epistemological divide. The various forms of participation discussed above are all possible and do not exclude each other. However, genuine modes of engagement and the space of "participation as justice, citizens' control, and empowerment" are particularly challenging as AI is developed in and for corporate or public-sector contexts, which are inherently extractive and oppressive. PAI builds on centuries of systemic and institutionalised injustice and inequity, especially in the technology development sector, and (in most cases) fails to acknowledge how AI development adds burdens on communities that consistently experience discrimination, stigmatisation, and inequity [103]. It remains difficult for designers and engineers conducting PAI to account for how historical mechanisms of oppression have shaped design spaces and practices [33]. As Birhane and colleagues [18] remind us, "*When we invite people to participate, it is never everyone*": the ones who are typically excluded, those with low literacy, little time, and living at the margins, never gain a seat at the PAI table, let alone a voice. This systematic exclusion is further exacerbated by the widespread belief, even among experts, that AI algorithms are too complex to be meaningfully explained, leaving non-experts without the instruments needed for proper deliberation [47, 113]. Feminist and decolonial AI critiques, however, warn us about the fallacies of this thinking and argue that the very problem lies with the dominant AI culture [23], what McQuillan ([90] page 110) provocatorily calls "*AI's epistemological apartheid*", which, under the mirage of objectivity and scientism, neglects certain ways of knowing grounded in the lived experiences of marginalized and minoritized groups. This epistemic bias leads to the embedding of dominant worldviews and social assumptions

into the very foundations of AI systems, shaping how we understand both the natural and social worlds [90]. It is imperative, then, for the critical AI community to find ways to challenge AI scientism, to dismantle public perceptions of AI as a magical entity [86, 113], and to empower marginalised voices to tell alternative stories of AI [113]. More than AI literacy and expertise, what is needed is a co-defined language and pace for PAI to thrive and genuinely empower communities. Researchers, designers and engineers should invest time and resources in building a relationship with the involved communities, where terms for consent [33] and modes of participation [18] are negotiated. Meaningful participation demands long-term presence in AI research, institutions and practices [103] as well as dedicated infrastructures, forms of organised bodies that situate power outside of the public institutions [33], maybe in the fashion of what McQuillian envisions as *people's councils* – a grouping of those affected by AI that form a directly democratic body to decide what to do about it ([90] page 127) – or more concretely, as what we are witnessing with bottom up tech workers initiatives, like *Tech Workers Coalition* and *African Tech Workers Rising*, just to name a couple.

2.2.3 PAI risks being performative. Despite the growing awareness of our responsibilities as contributors to AI development, most of the PAI efforts risk being *performative* [33, 114]. Stakeholders tend to be involved for providing input on discrete implementation parameters, rather than to be empowered to make key AI design decisions [36]. Experts very often do not even take the needs or recommendations of the different stakeholder groups into consideration [114]. Participants' engagement often serves as an aid in the refinement of AI systems; efforts are capitalised by corporate actors to maximise profit [18], and participation becomes once again a form of corporate co-optation [18].

Many of these issues stem from the capitalist infrastructure that AI "lives in" and the systemic problems of power and oppression that our society suffers from. Yet, we as researchers, designers and engineers practising PAI have a role to play; we can and must strive for justice when performing PAI. The problem, however, is that even when knowledge is present, this may be hard to access and even harder to put into action. In recent years, scholars have helped clarify the core dimensions of participation, underscoring its purpose, identifying who it should serve, and exploring how it can be meaningfully applied in AI contexts [18]. Literature provides us with ways to compare and contrast how different approaches can affect the power afforded to people. However, significant variations still exist in the methods and theories used to achieve participation [36], and the structural conditions these operate within often remain overlooked. Thereafter, a thorough scrutiny of how current PAI practices relate to social justice ambitions is needed, if we are to build a PAI culture that truly strives to contribute to the development of AI systems that are beneficial to society.

3 A critical scrutiny of Participatory AI through the lens of Design Justice

In this work, we examine how PAI practices engage with questions of social justice by conducting a structured scoping literature review of articles published in leading HCI venues. Our goal is to highlight and critically reflect on contributions that adopt explicitly critical perspectives on participation in AI, with particular attention to equity, power, and inclusion. In conducting our scoping literature review [6], we selected top-tier HCI journals and conferences as data sources and identified relevant studies through targeted keyword searches. We then performed a deductive thematic analysis guided by a collaboratively developed codebook. Grounded in Design Justice Principles, the codebook was iteratively refined through team discussion and annotation. This process enabled us to systematically analyse each article across six thematic dimensions, reflecting the *why*, *what*, *who*, *how*, and *when* of participation.

3.1 Design Justice as a lens for critical inquiry

The concept of justice can take many forms and articulations, depending on the practical context in which it is addressed, and the disciplinary lens it is looked at from [91]. To better articulate which configuration of justice we engage with and why, we provide here a brief discussion of current theoretical lenses that the critical AI community engages with, and how our work sits within this landscape. Current literature from philosophy, political science, sociology, STS, and other fields addresses these social justice issues through a variety of theoretical lenses. For instance, the *Capability Approach* –a philosophical framework that evaluates social justice by focusing on people's real opportunities to live a dignified life [106]– is often used to underscore that when AI is used for important functions of our society, like education and healthcare, one's capability to act upon, redirect and even reject specific dinamics may be limited, especially when not everybody has equal access to AI technologies [25]. In this regard, some scholars argue for looking at the relationship between AI and justice as a matter of *fairness* [136]: the idea that everyone has equal basic liberties compatible with the same for all; social and economic inequalities are allowed only when tied to positions open to all under fair equality of opportunity and arranged to maximise the benefit of the least advantaged [105]. In this vein, technical approaches to algorithmic fairness have grown in popularity. However, these are increasingly criticised for their incapacity to take into account the context in which these systems are applied [49, 73], and the network of socio-technical relations they are entangled in [74, 111]. As technical approaches fail to "account for the mechanisms by which our non-ideal world arises" [49], social sciences and STS scholars increasingly call for alternative epistemologies that look at the complex entanglement of AI with our social world. In particular, feminist, queer, and black philosophies expand the current discourse around AI ethics to underscore how AI-related issues, our capability of action, and the possibility of achieving fair systems are intrinsically intertwined with power relations [110]. AI manifests our societal power structures in a multitude of ways, spanning from economic dominance, uneven distribution of access to resources, asymmetries in decision-making, and discrimination [118]. Black feminist theories, in particular, stress how problems associated with power and the related systems of oppression, such as racism, sexism, and classism, do not exist in isolation, but are rather *intersectional* and create overlapping inequalities [31]. This also means that "*marginalised and subordinated communities find themselves facing social problems that can neither be understood nor solved in isolation*" [32]. Thereafter, these alternative perspectives

argue for AI to be developed in fairer, slower, consensual, and collaborative ways, and to do so by including diverse actors in the making of AI as well as involving communities in small-scale, slow data and AI projects [130].

Building on these perspectives, in which justice is collectively constructed and fought for, design scholarship has increasingly defined how the discipline should account for matters of social justice in practice. In particular, a community–the *Design Justice Network*– distilled a set of ten justice-informed design principles (1) that underscore the essential role of enabling communities to have voice and agency in technology development [40]. We consider the Design Justice principles as a potentially powerful instrument for critical inquiry in the space of PAI, as these build on the theoretical foundations of feminist, queer and black epistemologies, but advance knowledge in terms of what it takes to put these theories into practice.

Thereafter, in this work, we use the Design Justice Principles [40] as an analytical lens that foregrounds the lived experience, leadership, and agency of communities most impacted by socio-technical systems. While these principles were not developed specifically for AI, they offer a compelling framework for interrogating how participation is conceptualised and enacted in technology design, particularly in relation to justice. Their focus on co-governance, redistribution of power, and recognition of collective knowledge made them especially well-suited for guiding our coding and interpretation of the literature.

3.2 Positionality

In line with feminist HCI scholarship [13], we acknowledge the interpretative and situated nature of our approach [65] and therefore articulate the positionalities that informed our investigation with the current landscape of critical AI research. As three Assistant Professors working across design, HCI, and AI, we position ourselves within feminist, critical, and justice-oriented traditions in technology research. Commitments to diversity, equity, inclusion, and the interrogation of power in sociotechnical systems shape our shared standpoint. One of us approaches this work through militant design and critical AI, motivated by both personal passion and the urgent need to resource participatory AI projects. Another grounds their scholarship in posthuman feminism and activist practice, developing transdisciplinary methods to reimagine AI and robotics alongside workers, unions, and civic actors. The third brings a design research perspective attentive to democracy, accountability, and the methodological infrastructures that govern participation in AI. Together, we seek to build reference points and frameworks that not only advance academic debates but also sustain collective practices of resistance, reappropriation, and reimagining in human-technology relations.

3.3 Identifying data sources and corpus selection

3.3.1 Databases, Conferences and Journals. The review was conducted in the Association for Computing Machinery Digital Library (**ACM DL Library**), as it indexes the primary venues in HCI and related fields where critical perspectives on AI are published.

We targeted conferences and journals with a demonstrated dedication to the matters of social justice, including the ACM Conference on Human Factors in Computing Systems (**CHI**), the ACM Designing Interactive Systems Conference (**DIS**), the AAAI/ACM Conference on Artificial Intelligence, Ethics, and Society (**AIES**), the ACM Conference on Equity and Access in Algorithms, Mechanisms, and Optimization (**EAAMO**), the ACM Conference on Fairness, Accountability, and Transparency (**FAACT**), the ACM Conference on Computer-Supported Cooperative Work & Social Computing (**CSCW**), the ACM/IEEE International Conference on Human-Robot Interaction (**HRI**), the ACM Participatory Design Conference (**PDC**), the ACM Transactions on Computer-Human Interaction Journal (**TOCHI**), and ACM Transactions on Human-Robot Interaction Journal (**THRI**).

3.3.2 Search strategy. The search terms were designed to capture research explicitly engaging with participatory and justice-oriented approaches in AI:

> **All: participatory ai] AND [Abstract: design] AND [Abstract: justice] AND [E-Publication Date:(01/01/2019 TO 31/07/2025)]**

Articles were included if containing the phrase '***participatory AI***' anywhere in the text, and '***design***' and '***justice***' in the abstract. The search covered publications from January 2019 to July 2025. We set January 2019 as the starting date for our corpus, as this corresponds to the period in which the EU High-Level Expert Group on AI presented the Ethics Guidelines for Trustworthy Artificial Intelligence [68], which marked a key milestone in the institutionalisation of participatory and ethical concerns in AI. The official presentation of the guidelines followed the publication of the first draft in December 2018, which received more than 500 comments through an open consultation. The initial search, conducted between May and July 2025, resulted in a corpus of 202 articles.

3.3.3 Exclusion and Inclusion Criteria. We limited the review to research articles as they are thoroughly peer-reviewed, excluding other formats such as demonstrations, late-breaking works, workshop proposals, student design competitions, and video presentations. In line with our disciplinary standpoint in design research and our commitment to a holistic understanding of participatory AI processes, we focused this investigation on participatory processes dedicated to the design and development of an AI system. We excluded studies that conducted participatory practices with the intent of understanding AI's impact on people and society, without providing clear insights and recommendations for designing AI systems. All three authors collaborated iteratively in the selection process. Titles and abstracts of the 202 articles were screened, with the corpus divided into three sub-corpora. We checked whether the work actually addressed the theme of participatory design, social justice, and the design of AI systems. The absence of at least one of these themes represented an exclusion criterion. This phase resulted in a reduced corpus of **65 articles**. After the three authors independently annotated, articles were further excluded if they appeared not to be related to AI systems. The resulting corpus comprised **26 articles**.

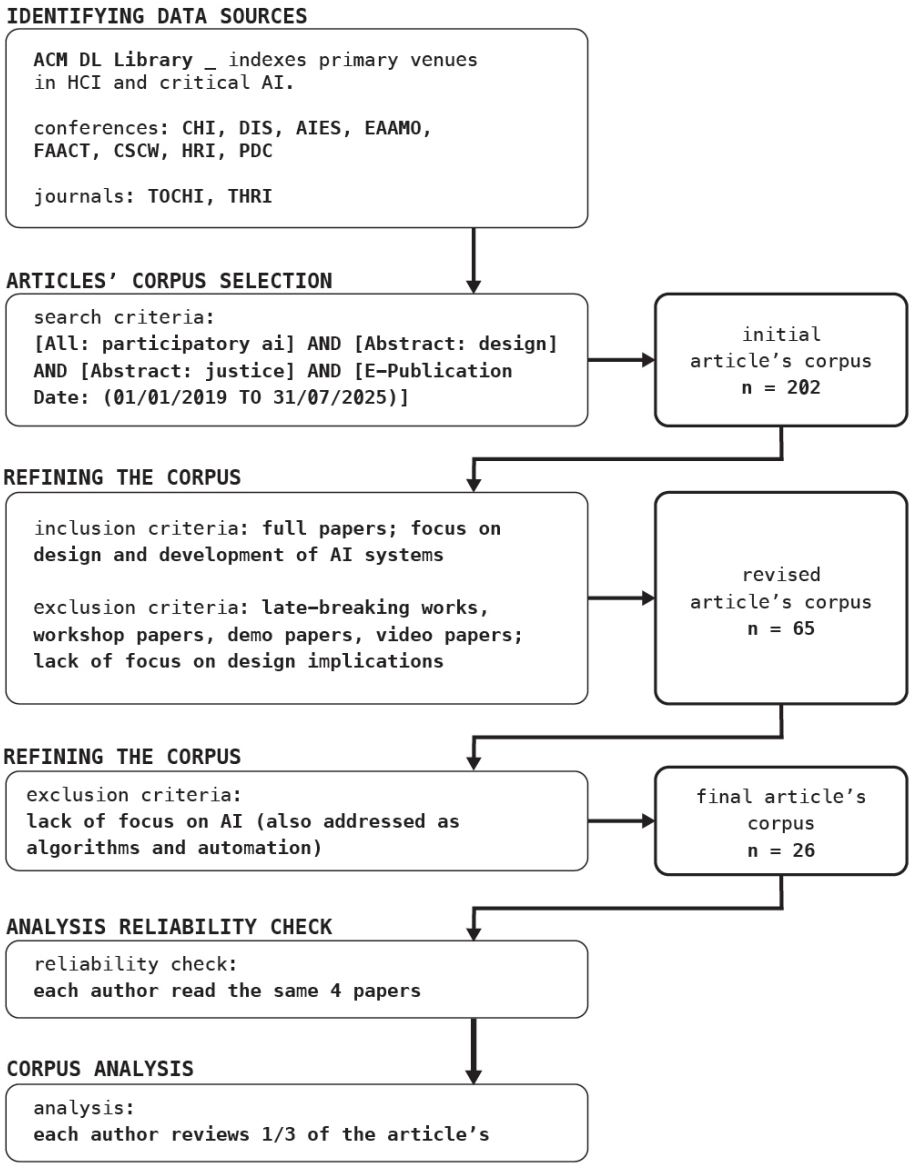


Figure 1: Overview of the scoping review process

3.4 Corpus and data analysis

We carried out a deductive thematic analysis [22, 100] where the deductive approach was chosen to enable systematic examination of how current scholarship incorporates existing critical perspectives and related designerly methodologies in participatory AI. This approach aligned with the overarching goal of the review, which was to map and assess how participation and commitments to social justice are conceptualised and operationalised within the field.

The theoretical background of our analysis is informed by feminist and decolonial critiques of AI [12, 24, 34, 57, 80, 82, 97], joining current efforts within the HCI community to articulate the limits of participatory AI. In particular, we share with Delgado et al. [36] and Corbett et al. [33] a commitment to examining how power is (or not) challenged and redistributed through participatory processes. Distinctively, however, our work addresses the implications of engaging with matters of social justice in participatory AI. Similar to Feffer et al. [50], we aim to provide a more holistic understanding of this practice space, extending beyond the focus on who is involved and how. For this, we structured our analysis around the Design Justice Principles [40] (Table 1), which encourage reflection not only on the subjects and modes of participation, but also on the very intentions behind design interventions, the role of 'designers', and the extent to which alternative ways of knowing are genuinely recognized and valued [35, 53].

Table 1: The *Design Justice Principles* [40] that guided the data analysis

Principle	Description
Principle 1	We use design to sustain, heal, and empower our communities, as well as to seek liberation from exploitative and oppressive systems
Principle 2	We centre the voices of those who are directly impacted by the outcomes of the design process
Principle 3	We prioritise design's impact on the community over the intentions of the designer
Principle 4	We view change as emergent from an accountable, accessible, and collaborative process, rather than as a point at the end of a process
Principle 5	We see the role of the designer as a facilitator rather than an expert
Principle 6	We believe that everyone is an expert based on their own lived experience, and that we all have unique and brilliant contributions to bring to a design process
Principle 7	We share design knowledge and tools with our communities
Principle 8	We work towards sustainable, community-led and -controlled outcomes
Principle 9	We work towards non-exploitative solutions that reconnect us to the earth and to each other
Principle 10	Before seeking new design solutions, we look for what is already working at the community level. We honour and uplift traditional, indigenous, and local knowledge and practices

3.5 Overview of the corpus

To contextualise the corpus and identify scholarly and geographic contexts relevant to power and epistemic dynamics [53, 83], we annotated each article by publication year, country of author affiliation, publication venue, and disciplinary background of the article. Publication year was recorded as reported in the article. The country of affiliation was derived from the institutional affiliation of the authors. Disciplinary background was coded using the **ACM Computing Classification System**.

3.6 Codebook development

The codebook was developed deductively, building on current critiques to PAI and participatory design, and articulated in the Design Justice Principles [40]. The three authors collectively discussed and grouped these principles into guiding themes, which were then translated into analytical questions for annotating the corpus. Each theme was defined with reference to relevant principles, accompanied by a set of annotation dimensions. The codebook was iteratively refined through repeated annotation and discussion until consensus was reached. The final codebook comprised five thematic foci, each associated with specific Design Justice Principles:

- **Why**. *What is the purpose/ambition of the participation work?* (relates to principles 1 and 9). This theme centres on the goal of the design practice and/or research. It looks at whether the participatory practices are oriented to support, sustain, empower and heal communities in a non-extractive way with a clear emancipatory or liberatory orientation. The goal of the design is non-exploitative and helps connection.
- **What**. *What is being designed? Is it a tool for the community? Is it AI? What kind of design work is the research trying to inform?* (relates to principles 8 and 10). This theme centres on what is being designed or produced through research, and the intended use of resulting knowledge or artefacts.
- **Who Involves**. *What is the role of the designer?* (relates to principle 5) This theme centres on who initiates participation, the role of designer(s) or researcher(s), and their awareness about power dynamics, situatedness of the participatory practices and the type of partnership established with the participants.
- **Who is involved**. *Which community is involved?* (relates to principle 2) This theme centres on which communities or participants are (or should be) involved in the design process and how they are described.
- **How**. *How does participation unfold?* (relates to principles 3, 4, 6 and 7) This theme centres on how participation is established and mediated by designed artefacts.

During iterative annotation to reach consensus over the corpus and codebook, we added an additional theme (reaching six in total) to account for the phase of the design process of the participatory practices. The resulting question and thematic description are as follows:

- **When**. *When in the AI lifecycle does the participation unfold or should unfold?* This theme takes into account the definition of Corbett et al. 2023 [33], considering the lifecycle process as comprising four interconnected steps, i.e., problem formulation, dataset development, model design and training, and deployment and monitoring.

3.7 Data analysis and reliability

To ensure the reliability of the themes and guiding questions, four randomly selected articles (14% of the final dataset) were independently screened, read, and annotated by all three authors. The resulting annotations were compared and discussed in detail to strengthen the reliability of the codebook. One article was subsequently excluded due to a lack of thematic relevance. We conducted several consensus discussion rounds and refined the questions after reading each of the three articles.

Following the codebook refinement and consensus discussions, the full dataset was divided evenly among three authors, with each coding one-third of the corpus (9 articles each). The data analysis was performed by reading the abstracts, introduction, methods

section, positionality statements (if available), discussions, and conclusions. When necessary, full texts were carefully read. On average, each article took about 25 minutes to review. The authors used the guiding questions and related definitions to annotate the articles and noted the instances of certain facets being present or not, or added answers to the guiding questions.

4 The corpus of articles

4.1 Year of publication

The majority of articles included in the corpus were published in 2023 (8 articles [37, 43, 51, 62, 72, 101, 116, 128]) and 2024 (8 articles [1, 8, 9, 63, 71, 76, 77, 127]). Three articles were published in each of the years 2021 [56, 98, 137], 2022 [45, 58, 114], and 2025 [81, 123, 126]. Only one article in the corpus was published in 2019 [135] and none in 2020.

4.2 Country of affiliation

The articles in our corpus predominantly report research conducted by researchers based in WEIRD countries, particularly the United States which accounts for 16 of the 26 articles [1, 37, 43, 45, 51, 58, 62, 72, 76, 81, 98, 114, 116, 126, 127, 137]. The United Kingdom (UK) follows with two articles [56, 77]. Other countries represented include Canada [101], and Sweden [71], each with one article in the corpus. Six articles resulted from international collaborations, namely two articles with researchers based in the USA and Canada [9, 63], one article with researchers based in the UK and Ireland [128], and one between the Netherlands and Italy [8]. The collaborations between the UK, Bangladesh and the USA [123], as well as between Singapore and the USA [135], represent the only articles in the corpus that include a research institute based in the Global South.

4.3 Venue of publication

Most of the articles in our corpus were published at the CHI conference (nine out of 26 articles [56, 62, 63, 76, 101, 116, 126, 135, 137]). This is followed by the TOCHI journal with four articles [51, 81, 127, 128], and DIS conference with another four articles [9, 37, 71, 123]. The AIES, EEAMO and FAccT conferences are represented in the corpus, each with three articles [1, 8, 43, 45, 58, 72, 77, 98, 114].

4.4 Disciplinary background of the article as per CSS Classification

There are four macro subsets of disciplinary backgrounds described by category and concepts of the computing field: (i) Computing Methodologies and Applied Computing, with four articles [43, 58, 98, 128]), (ii) Human-centred computing, with eighteen articles [9, 37, 45, 62, 71, 76, 81, 127, 137], (iii) Social and Professional topics, with three articles [1, 116, 126] and HCI design and evaluation methods, with one article [101]. Notably, one article does not report any CSS classification [8]. The Computing Methodologies and Applied Computing articles are further classified as focusing on a specific technical application (i.e., machine learning, natural language processing, law applied computing, 3 articles [58, 98, 128]) or they are related to computing philosophy and theory (1 article [43]). The Human-centred computing articles have various subclassifications, but they are predominantly classified as field and empirical studies and HCI theory. The Social and professional topics have various subclassifications, including education, government policy, and race and ethnicity. The HCI design and evaluation methods article present in the corpus is the second one to make an explicit participatory design classification [101].

5 Findings

5.1 Why do researchers/practitioners do the work of participation?

Justice-oriented PAI work aspires to develop more responsible AI systems and empower communities to steer AI innovation processes

Looking at '*why*' practitioners and researchers engage in justice-oriented PAI reveals a wide range of ambitions that we group in three overarching purposes: 1) *Designing for justice*; 2) *Developing responsible AI*, and 3) *Empowering communities*. *Designing for justice* is devised in 12 out of 26 articles [1, 45, 51, 56, 62, 71, 77, 98, 114, 123, 127, 137] and manifests as a strong and explicit orientation towards tackling power structures, inequality and discrimination. It typically proposes alternatives to the status quo through frameworks, methods or co-design practices. The articles emphasise celebrating and valuing the lived experience of communities and sharing agency in the design processes. We find 9 out of 26 articles [8, 37, 58, 63, 72, 76, 101, 128, 135] committed to *Developing responsible AI*, such as for co-developing an explainable diagnostic AI framework with clinicians [135] or co-designing a toolkit to support democratic decision making around public AI systems [76]. These mostly unfold through participatory engagements, often also oriented towards addressing matters of explainability, accountability, fairness, transparency, and alignment with socio-ethical and legal values. Unlike the justice-oriented works, these articles do not always foreground lived experience but instead position communities or stakeholders as contributors to making AI more accountable. A minority of works (4 out of 26 articles [9, 81, 116, 126]), focus on *Empowering communities*, namely directly engaging with marginalised communities or underrepresented groups, creating opportunities for people to exercise control and agency in AI design processes. For instance, Tanksley et al. [126] focus on enabling Black Youth to reimagine AI technologies, and Koo et al. [81] aim to give workers greater power over technological mediation in their workplaces. Works in this category differ from *Designing for justice* in the extent of engagement. The commitment to marginalised communities and alternative epistemologies is shared, yet *Designing for justice* attempts to span the whole design process, while the *Empowering communities* occurs in specific moments of the process.

Nevertheless, the works in the corpus tend to share an overarching ambition to develop more responsible AI systems, alongside a commitment to empowering communities to have a role in steering AI innovation processes, including through actions aimed at dismantling power structures, inequality and discrimination as a pathway toward design for justice.

Table 2: The *Articles included in the final corpus*

Author(s)	Title	Venue	Year
Petterson et al.	Playing with Power Tools: Design Toolkits and the Framing of Equity	CHI	2023
Abdu et al.	Algorithmic Transparency and Participation through the Handoff Lens: Lessons Learned from the U.S. Census Bureau's Adoption of Differential Privacy	FAccT	2024
Thieme et al.	Designing Human-centered AI for Mental Health: Developing Clinically Relevant Applications for Online CBT Treatment	TOCHI	2023
Kawakami et al.	The Situate AI Guidebook: Co-Designing a Toolkit to Support Multi-Stakeholder, Early-stage Deliberations Around Public Sector AI Proposals	CHI	2024
Arzberger et al.	Nothing Comes Without Its World – Practical Challenges of Aligning LLMs to Situated Human Values through RLHF	AIES	2024
Asha et al.	Shotitwo First!": Unravelling Global South Women's Challenges in Public Transport to Inform Autonomous Vehicle Design	DIS	2024
Gansky & McDonald	CounterFAccTual: How FAccT Undermines Its Organizing Principles	FAccT	2022
Halperin et al.	Probing a Community-Based Conversational Storytelling Agent to Document Digital Stories of Housing Insecurity	CHI	2023
Whitney et al.	HCI Tactics for Politics from Below: Meeting the Challenges of Smart Cities	CHI	2021
Wang et al.	Designing Theory-Driven User-Centric Explainable AI	CHI	2019
Edenberg & Wood	Disambiguating Algorithmic Bias: From Neutrality to Justice	AIES	2023
Johnson & Crivellaro	Opening Research Commissioning To Civic Participation: Creating A Community Panel To Review The Social Impact of HCI Research Proposals	CHI	2021
Solyst et al.	"I Would Like to Design": Black Girls Analyzing and Ideating Fair and Accountable AI	CHI	2023
Taylor & Bruckman	Mitigating Epistemic Injustice: The Online Construction of a Bisexual Culture.	TOCHI	2024
Nee et al.	Advancing social justice through linguistic justice: Strategies for building equity fluent NLP technology.	EEAMO	2021
Kay et al.	Epistemic injustice in generative AI.	AIES	2024
Forlano & Halpern	Speculative Histories, Just Futures: From Counterfactual Artifacts to Counterfactual Actions	TOCHI	2023
Haque et al.	Are We Asking the Right Questions?: Designing for Community Stakeholders' Interactions with AI in Policing	CHI	2024
Sloane et al.	Participation Is not a Design Fix for Machine Learning	EEAMO	2022
Jääskeläinen et al.	AI Art for Self-Interest or Common Good? Uncovering Value Tensions in Artists' Imaginaries of AI Technologies	DIS	2024
Sultana et al.	'Socheton': A Culturally Appropriate AI Tool to Support Reproductive Wellbeing	DIS	2025
Dillahunt et al.	Eliciting Alternative Economic Futures with Working-Class Detroiters: Centering Afrofuturism in Speculative Design	DIS	2023
Tanksley et al.	Ethics is not neutral": Understanding Ethical and Responsible AI Design from the Lenses of Black Youth	CHI	2025
Jegede et al.	Challenge Accepted? A Critique of the 2021 National Institute of Justice Recidivism Forecasting Challenge	EEAMO	2023
Ehsan et al.	The Algorithmic Imprint	FAccT	2022
Koo et al.	Metrics and Macchiatos: Challenges for Service-Industry Workers and the Need for Worker-Driven ICTs	TOCHI	2025

5.2 Who involves

PAI researchers/practitioners show a commitment to community engagement, yet they rarely reflect upon their position and power

Several articles offer theoretical or methodological frameworks without explicitly addressing '*who involves*' participants in the PAI process. It is generally understood that researchers and practitioners occupy this role in PAI work, whether as facilitators, enablers of co-creation, or experts guiding the studies. However, the ways in which they take up this responsibility vary significantly, ranging from approaches that maintain clear boundaries between "researcher" and "participant" to those that cultivate more symmetrical partnerships where these roles blur.

These differences become clearer when looking at how specific studies describe researcher roles in practice. Four articles [9, 37, 62, 137] illustrate cases in which the conventional distinctions between the researcher and participant are softened. For example, Whitney et al. [137] and Halperin et al. [62] describe their immersion in the communities they worked with, e.g. by serving on city coalitions, participating in steering committees, or building on long-standing relationships rooted in sustained community engagement. Asha et al. [9], meanwhile, demonstrate a different form of boundary building: Rather than engaging with a separate participant community, the researchers draw on their first-person experiences as women from the Global South navigating public transportation. In this case, the researchers themselves are the community whose lived experience becomes the basis of inquiry.

In contrast, other studies portray a clearer differentiation between researcher and participant roles. Two articles [116, 126] describe researchers acting as teachers who involve their students in participatory activities that combine learning with critical engagement. Another article [76] reports on the use of professional facilitators from a private company, further reinforcing a division of roles by outsourcing facilitation to specialists whose professional function is separate from both researchers and participants. These examples show instances where responsibility for involving participants is more formally assigned and less interwined with researchers' own identities or lived experiences.

Despite these diverse configurations of involvement, only eight out of 26 articles in the corpus include a positionality statement [9, 37, 71, 81, 116, 123, 126, 127]. These are primarily empirical or design-oriented studies that reflect on how the researchers' identities impacted the research. For example, Taskley et al. [126] acknowledge that "*Our identities, scholarship, and positionalities shape the lens through which we analyse these systems, allowing us to critically engage with AI's role in exacerbating existing inequities and to offer frameworks that centre justice and equity*". Koo et al. [81] provide a detailed account of how relationships between researchers and service workers are deeply entangled with broader labour relations. Taken together, despite these works' commitment to involve communities, power dynamics connected to the identity, background, position and intentions of the researchers/practitioners, remain under-addressed.

5.3 Who is involved

PAI participants range from expert stakeholders to community members, but seldom from community-led organisations

Examining the corpus through the lens of '*who is involved*' reveals a wide diversity of communities, experts, and stakeholders involved in PAI projects. Several articles work directly with specific communities embedded in particular social contexts, for example, Black girls critically reflecting on AI accountability [116], and women in the Global South addressing safety in public transportation [9]. Other studies centre professional experts in AI and various other fields in the participatory process. For instance, Wang et al. [135] worked with clinicians to co-design explainable AI, while Haque et al. [63] involved technical specialists and policy experts. These projects highlight the importance of domain expertise for shaping responsible and context-sensitive AI systems. Several articles with a more theoretical or methodological orientation do not involve participants directly but instead call for the engagement of broader communities in participatory AI practices. For example, Abdu et al. [1] and Sloane et al. [114] advocate for expanding the scope of participation to a broader array of "stakeholders", while Nee et al. [98] specifically call for the participation of diverse (unspecified) language communities in the design of natural language processing technologies.

Across these works, community members, field experts, and other stakeholders play visible roles in participatory processes. However, *participants rarely come from community-led organisations*, despite their potential to shape AI development in ways that center collective agency and power. A notable example is the work by [137] that reports on bottom-up technological practices instantiated by groups of residents organising to demand a role in the design and oversight of urban computational technologies in San Diego (US). As this work emphasises, compared to the involvement of stakeholders and individual community members, engaging with community-led organisations is an essential practice to enable truly just processes, as these allow for genuine contestation, refusal and negotiation of power.

5.4 What is being designed?

Justice-oriented PAI prioritises knowledge production and social change over the development of AI artefacts

Analysing the corpus to understand '*what is designed*' reveals two broad categories of outcomes: *Design Possibilities* and *Theory and Methods*. Seven articles [51, 56, 71, 116, 126, 135, 137] involve co-creation, speculative design, or co-prototyping activities that produce alternative imaginaries and design directions as *Design Possibilities* that stand in contrast to mainstream AI innovation. These studies frequently use artefacts such as mockups [135] and canvases [37, 71, 137], not as final products but as prompts for reflection, critical discussion, and collective envisioning. Ten articles [1, 8, 43, 45, 58, 72, 77, 98, 101, 114] focus instead on producing

Theory and Methods. These present critical essays, position papers, and theoretical or methodological frameworks that offer conceptual tools for addressing PAI. For example, Abdu et al. [1] present a framework to make visible how values are embedded within sociotechnical AI systems and to configure who participates in their design. Nee et al. [98], instead, introduce a framework to help NLP researchers engage with matters of linguistic justice.

A further eight papers [1, 43, 71, 72, 76, 77, 101, 128] do not explicitly centre a design artefact or method, but nevertheless make important contributions by analysing critical dimensions of the design space. These include articulating the sociopolitical conditions under which PAI unfolds, identifying barriers to participation, or outlining constraints and tensions that shape how communities engage with AI systems. Such contributions help situate PAI within broader contexts and clarify the challenges that design work must address.

Across these categories,, the object of design is often intangible and conceptual, rather than a prototype—an AI thing. The corpus overall underscores the mostly immaterial nature of design outcomes, where knowledge and social change are prioritised in the face of AI product developments. As Forlano & Halpern [51] argue: "*As technology continues to play an essential, and even an existential role in the future of society, it is vital to continue to find ways of critically engaging the public with the ethical and political stakes around these choices*". In this context, social change frequently involves challenging prevailing norms or dominant practices (as illustrated in [56]) and creating spaces for communities to confront the wider consequences of PAI interventions. At the same time, several authors point out that participation itself can impose demands on community members—such as additional workload, limited resources, or emotional strain—which must be recognised and handled with care [123].

5.5 How does participation unfold?

PAI makes extensive use of designed artefacts, yet these function as instruments of participation rather than primary outcomes

With regard to the '*how*' PAI work is practised, the articles show a spectrum of approaches. The majority [9, 43, 45, 56, 62, 63, 72, 76, 81, 123] describe an *informing* relationship, in which researchers and participants exchange information to guide the design of AI systems. Several articles advocate for, or attempt to implement, *equitable partnerships* [37, 51, 101, 126, 137] where power and agency are shared, and reflexivity is encouraged. Other works adopt more *evaluative* modes of involvement, such as tester roles where participants help assess technologies [128], or consulting partnerships, where non-researchers provide targeted feedback [77, 127, 135]. Further, the majority of the articles recognise the value of lived experience and centre these in the design or sensemaking processes. Engagement with lived experience is seen as an immersion of reflexivity and a continuous productive confrontation [137], a key process for improving the design of AI systems sensitive to the life experience of marginalised groups [43].

Supporting these various modes of participation approaches, the articles present a rich engagement with design materials [1, 8, 9, 51, 58, 62, 71, 76, 77, 98, 101, 114, 116, 123, 126, 135, 137] that vary from tools, toolkits, prototypes and other objects to mediate participation. This mediating role of artefacts is particularly evident in a public-engagement context, as in the work by Dillahunt et al. [37] who use a speculative Afrofuturist workbook—*Building Utopias workbook*—to probe participation and to encourage speculative and critical design thinking from an Afrofuturism lens in PAI processes. In PAI work, then, artefacts or pieces of technology become **boundary objects** [119]: artefacts that are robust enough to maintain a common identity across different social or disciplinary worlds, yet flexible enough to be interpreted locally by each group [119]. However, as noted by Abdu et al. [1] "*The mobilisation of any given boundary object is dependent not only upon the object itself, but also upon the motivation and orientation of those brokers that span and connect communities*". Thereafter, the very materiality of AI is considered by authors a way to engage with the sense-making of participation [77]: material engagement with the AI mediate the experience "*AI can be an interactive tool for exploring one's own experiences. Image generation can be used to re-imagine and express oneself*". These works, then, engage with tools, toolkits, mockups and other artefacts to facilitate participation, but the focus remains on the procedural aspects of PAI work not the design of the artefacts themselves.

5.6 When does participation unfold?

The majority of PAI work focuses on the early phases of the AI development cycle

The majority of articles [1, 9, 37, 43, 51, 56, 62, 71, 72, 76, 77, 81, 98, 116, 123, 126, 127, 137] focus on the *problem definition* phase, while the rest of the corpus focuses on participation unfolding in the *deployment* and *monitoring phase* [45, 63, 128, 135], and the *complete lifecycle* [8, 58, 101, 114].

Research articles that highlight the importance of community involvement in the problem definition activities argue that focusing on this early phase of development is crucial for addressing socio-technical issues before actual technology development begins. Edenberg and Wood [43], for instance, call us to "*redirect approaches to addressing bias in generative AI at its early stages in ways that can more robustly meet the demands of justice*" (p.691). These works focused on the problem definition phase are primarily oriented to understanding people and communities in context and include many processes and methods such as surveying communities [1, 81, 123, 127], collective ideation and prototyping [56], the use of mock-ups [128] and speculative sessions [37, 51, 126].

In the deployment and monitoring phase, participation unfolds to understand possible opportunities for re-design as well as to understand the implications of deployment. Ehsan et al. [45] introduce the notion of "algorithmic imprinting" arguing that the consequences of algorithms extend further than the deployment period and even when an algorithm is eventually removed from a system and more attention should be given to situated fairness:

"*Awareness and traceability of the contours of the imprint can facilitate improved active participation in algorithmic mediation. An imprint-aware design mindset treats stakeholders as active (as opposed to passive) participants in the design process*" (p. 1312 [45]).

Interestingly, of the articles focusing on the complete lifecycle or advocating for participation unfolding in the full process, a few provide interesting insights about current limitations related to "when participation unfolds". Gansky and McDonald [58], for instance, highlight how there is a tendency to "*focus on a single stage of development at a time, with the intention of affecting conditions across the entire life cycle*" and argue for more attention to be paid to the socio-technical and organisational context of AI deployment. In this regard, Ehsan et. al [45] call for extending the locus of analysis and looking at the junctions between use (lifecycle) and post-use (afterlife), for a more holistic assessment of algorithmic impact. In support of this, Petterson et al. [101] suggest that toolkits hold potential to support more holistic participation practices in extended AI lifecycles.

6 Discussion

Our analysis reveals that practising justice-oriented PAI work remains more an ambitious than a practical reality, as many of the articles present theoretical frameworks instead of empirical work. Nevertheless, there are several instances of how virtuous practices are instantiated and important reflections on our role and approaches are encouraged. In what follows, we provide further reflections on these methodological implications. We particularly focus on the role of the PAI researcher/practitioner and the importance of building alliances with communities, along with a reflection on the important, yet not self-sufficient, role of artefacts in PAI processes.

6.1 Practising PAI, being an activist

What becomes clear by looking at virtuous examples doing the work of social justice in participatory AI is that the very identity of the PAI researchers and practitioners is a hybrid one. Powerful and holistic PAI work, in fact, is most often done by people who fit both the role of the researcher/practitioner and the one of participants in the research (or at least a member of the community that the research looks at). For example, Koo and colleagues' work [81] about the pressure of workplace technologies on service-industry workers stems from the personal experience of the second author working for two years as a barista at Starbucks and the first author's experience as a member of the United Campus Workers union. The engagement of participants in this study heavily and openly relies on the personal networks of the authors. In the same vein, the work by Whitney and colleagues [137] on understanding and influencing the design and implementation of smart streetlight infrastructures in the city of San Diego is strongly motivated by the last author being a coalition steering committee member. Relatedly, the work by Asha and colleagues [9] on the potential of autonomous vehicle technologies for women in the Global South is motivated by the daily personal mobility experience and struggles of the first two authors, who were born and raised in Bangladesh, and identify as women. Personal experience and interests of the PAI researchers/practitioners, then, represent not only a motivator for engaging with specific communities, but also a facilitator, as this allows for leveraging networks, eases communication, and may reduce hostility and mistrust.

When the "personal" is at stake, **PAI research becomes a form of activism**. This aligns with the heritage of design tradition and its commitment to socio-political matters. As Bieling [17] articulates in their book "Design (&) Activism: Perspectives on Design as Activism and Activism as Design", over the last decades, the increasing vigour, diversity and number of popular and social movements has encouraged an extensive use of designerly methods and tools in support of individuals and communities to help expressing themselves and having a voice in matters of public concerns. Design activism clearly has a strong tie to political activism, but also materializes in distinct acts: instead of boycotts, strikes, protests, or demonstrations, design activism "*lends its power of resistance by intervening in people's lives*" [88]––it focuses on developing artifacts that exist in real time and space, and are situated within everyday contexts and processes of social and economic life [75]. PAI as activism blends these two perspectives as it both instantiates forms of public aggregation and contestation, as well as makes use of artefacts to contest dominant socio-technical paradigms and facilitate the collective shaping of socially desirable alternatives. Design workshops and hackathons become sites where PAI researchers/practitioners join forces with community organisations and members to voice dissent and envision alternative socio-technical paradigms [37, 71, 76, 137], in which artefacts and mockups are designed and used to support dialogue [1, 62]–to function as *AI boundary objects*.

This hybrid identity of the PAI researcher/practitioner is a key enabler of meaningful engagement with communities. Still, it is also a **hard-to-maintain** and an **impossible-to-prescribe** dimension of these practices. As Harrington and colleagues [64] argue, in fact, equity-driven and community-based participatory design practices are well-accepted both inside and outside of academia, yet giving them continuity is hard, especially after fundings end. A different academic culture where alternative forms of impact are acknowledged is needed [64], but also the toll that these practices impose on individuals, both in terms of time [18] and emotional labour [64], must be understood. Once again, design scholarship underscores that there are limits to the affordability of critical practice, as designers (and by extension PAI researchers/practitioners) also have to work under capitalist conditions [89]. It is important to acknowledge that one can act from a critical standpoint, in radical and activist terms, at one precise moment in their career, but not later, because circumstances can change, both in the personal sphere and the disciplinary discourse [85]. But, *what happens when a PAI researcher/practitioner engagement with a community ends?*

If we are to claim the merits of justice in our work, it is also essential to account for the effects that our interventions have on communities and contexts, beyond the exact period of our intervention. We should avoid what we can metaphorically name '*pride parade*' engagements– acts of performative allyship that are showcased only at the time of an event [133]– and account for the consequences of us both entering and exiting a community. In 'traditional' design processes, the value and impact of a participatory design intervention can be assessed in the way a product, service or system that has been collectively developed, gets actually adopted and used [21]. When it comes to practising PAI with the ambition

to fight for social justice, the very engagements, the gatherings, and the resulting forms of **contestation and resistance** constitute the value of the work [33, 59, 129]. But, *how can this be sustained over time?* Useful in this regard is to rethink once again the relationship between who leads PAI work and who participates. Instead of approaching these practices as individuals or teams of sole academics, **we better build large alliances with diverse actors** [115]. Dillahunt and colleagues [37], for instance, report on the collective exploration of alternative economic futures with Black and Brown working-class Detroiters. The team in their research is uniquely composed of academic researchers, community leaders and activists. This alliance required an extensive investment of effort and time to define shared goals and ways of working together, but holds the potential to provide continuity to the cause, as it doesn't rely on the availability and individual conditions of single subjects. For alliances to work, however, it is of utmost importance to understand the organisational dynamics of collectives that are "*shaped and reshaped in changing networks, co-operations, more or less temporary organisational synergies*" [115] and to be open to **rethinking the identity of PAI researchers/practitioners from individuals to groups**.

6.2 Practising PAI, crafting and using AI boundary objects

Central to many of the works reviewed here is an effort to provide the field with theoretical instruments to structure PAI work and/or artefacts for facilitating engagement of participants and their deliberation (see subsection 5.5). In both cases, we argue, the work can be framed as a practice of **crafting and using AI boundary objects**. This framing builds on the popular social science notion of *boundary objects*: artefacts that are robust enough to maintain a common identity across different social or disciplinary worlds, yet flexible enough to be interpreted locally by each group [119]. These are things that can be abstract or physical and "*exist in the liminal spaces between adjacent communities of people and have the capacity to traverse perceptual and practical differences among these and facilitate cooperation by fostering mutual understanding*" [69]. Because of this focus on the capacity to speak to different communities of practice and knowledge, the notion of boundary objects has established its role in innovation processes to enhance the capacity of an idea, theory, or practice to translate across culturally defined boundaries [52]. Within the space of participatory AI committed to social justice, boundary objects are potentially of crucial importance for their capacity to help understand, negotiate and break epistemic divides between researchers/practitioners and participants. AI boundary objects can help us challenge the "*AI's epistemological apartheid*" ([90] page 110) and the mirage of objectivity and scientism, creating bridges to alternative ways of knowing that are rooted in the lived experience of people who are marginalised or minoritised ([90] page 105).

A quick search of academic articles on AI boundary objects outputs works mentioning machine learning explanations [11], structural causal models [67], and onboarding materials [26] as boundary objects, just to name a few. This unveils a practical and technical view on AI boundary objects, where epistemic divides are only partially understood and addressed. Within our corpus, we find a few examples that align with this approach. Thieme and colleagues [128], for example, made use of mockups and their design rationale as a way to engage with experts to discuss clinically relevant AI applications, and the implications of design for AI applications in sensitive use contexts. Many of the reviewed works, however, make use of AI boundary *objects as artefacts materialising AI possibilities and to facilitate engagement in PAI work*. These do not necessarily engage with technical matters, but rather value lived experiences over expert knowledge. In some cases, the AI boundary objects are themselves objects of collaborative design and become sites where the boundaries between the PAI researchers/practitioners and participants get blurred. Jääskeläinen and colleagues [71] structured their participatory activities as a blend of collaborative ideation sessions, during which each participant was invited to imagine an AI tool for their creative practice, and a collective discussion aimed at raising critical reflections on the imagined possibilities. Here, the PAI researchers/practitioners did come with pre-made artefacts, i.e., *Speculative Sketching Templates*, but the very role of the AI boundary object was performed by the ideas represented on those templates. In a similar vein, Dillahunt and colleagues [37] provided participants with the *Building Utopias* workbooks and a card deck. These materials introduced new design and technology terms and served as a way for participants to both document their thoughts and collectively discuss. The work of Whitney and colleagues [137] is similar: they collaboratively designed artefacts with participants, during what they called a *Slightly Dystopian hackathon*. The work focused on developing speculative mockups of possible applications, building on the APIs of smart streetlights in the city of San Diego. Distinctively, however, their work emphasises how the mockups were not much valued for the ideas they were embedding, but rather for the capacity of these to help the coalition working on the initiative to problematize a 'broken' public use of data and engage the community around the problem.

We also find examples of frameworks and theories that can help us structure PAI work and that can potentially be used by teams to reflect on their PAI study designs. Many of these, such as the ones by Sloane and colleagues [114] and Jegede and colleagues [72], propose frameworks for better understanding participation in AI development and deployment processes, and potential issues. In a similar vein, Abdu and colleagues [1] propose the concept of *handoff* as a critical tool for untangling the daunting knot of actors, components, modes, functions, and values that should be accounted for in collective processes of deliberation focused on algorithmic innovation. What this type of work underscores is the possibly dual nature of AI boundary objects. While on the one hand, AI boundary objects can be used to ease technology and innovation adoption, on the other, these can also have the contrary effect, that is, to inhibit the uptake of a certain technology [52]. **AI boundary objects can function as a form of '*adversarial design*'** [38]: by manifesting controversial aspects of a technology, it opens up space for contestation and dissent–agoras for constructive agonism. As such, AI boundary objects hold great potential for practising PAI work that is committed to social justice: not only to achieve a more balanced distribution of knowledge between different subjects in participatory settings, but also to allow for agonistic feelings and opinions to be expressed.

Making AI boundary objects 'work', however, it's all but trivial. As Abdu and colleagues [1] articulate by reflecting on their study of the U.S. Census Bureau's adoption of differential privacy [2], the Bureau employed several conceptual boundary objects (i.e., interactive notebooks, webinars, handbooks, and videos) that resulted to be only partially successful. According to the authors, the causes of such limited success lies in an insufficient comprehension of the diverse ecosystem of stakeholders and a lack of connection with trusted local experts that could have carried the boundary objects across community divides. What the authors further underscore is that "*boundary objects cannot travel alone*" and it is crucial for the critical AI community to "*think about the expertise needed to shepherd and use such boundary objects effectively in order to broker meaningful trust and participation*" [1].

The craft and use of AI boundary objects in participatory AI, then, holds great potential for engaging with matters of social justice, and especially to open up spaces for agonism. However, it also comes with a multitude of disciplinary challenges. As PAI researchers/practitioners, we should ask what constitutes a good AI boundary object and what type of artefacts can function better to address specific aspects and implications of AI implementations. Along with such a focus on artefacts, however, **we should also instantiate models of engagement in which the craft and use of AI boundary objects is not a sole prerogative of PAI researchers/practitioners, but rather an instrument of the broader PAI coalitions** we mentioned in the previous paragraphs. We should ultimately investigate–collectively–what the cultural and contextual conditions are that an AI boundary object should fit in and how to better speak to these through things.

7 Conclusions

This work draws a picture of and reflects upon the current landscape of PAI research and practice in HCI, with a particular focus on matters of social justice and whether these are addressed through design, and in what forms. Grounding of feminist and decolonial AI critiques, and guided by design justice principles, we highlight current best practices, identify opportunities for future methodological expansions, and take stock of existing challenges.

Our scoping review of 26 articles from HCI venues finds that:

- Justice-oriented PAI work aspires to develop more responsible AI systems and to empower communities to stir AI innovation processes (**why**);
- The PAI researcher/practitioner (**who involves**) shows a commitment to involve communities, yet their position and power is rarely reflected upon;
- PAI participants (**who is involved**) are predominantly expert stakeholders or community members, but they rarely belong to community-led organisations;
- Justice-oriented PAI work prioritises the production of knowledge and social change to the development of AI artefacts (**what**);
- PAI work makes extensive use of designed artefacts, yet these are instrumental to PAI processes, not their outcomes (**how**);
- The majority of PAI work focuses on the early phases of the AI development cycle (**when**);

We further reflect critically on these findings and observe that there is a necessary methodological broadening that comes when practising PAI work, which concerns the blurry boundaries between 'who involves' and 'who is involved'. Our observation, grounded on a few virtuous cases (e.g., [37, 115]), is that instead of structuring PAI work as an endeavour that is exclusively led by the researchers/practitioners, **larger and long-term alliances should be built, especially with community-led initiatives and activist groups**. This is not only essential for facilitating community engagement, but also for ensuring continuity and impact of the PAI initiatives beyond the times and resources that are available to the individual researchers/practitioners. And along with encouraging a reflection on our very role as design and HCI researchers, we also suggest **rethinking once again the role that artefacts play within PAI processes**. Our observation here is that artefacts hold enormous potential for facilitating justice-oriented PAI work, especially if used as *AI boundary objects*. As such, design and HCI researchers/practitioners have an important role to play. Once again, however, **we must carefully investigate, collectively, the cultural and contextual conditions that an AI boundary object should fit in, and how PAI alliances can appropriate and make use of them**.

To conclude, this work contributes to critical AI and, more broadly, to HCI and design scholarship concerned with matters of social justice in participatory processes. To our knowledge, this is the first scoping review specifically focusing on PAI research in HCI explicitly analysed through the lens of social justice, and design justice more specifically. The overview we provide on the current state of the art of justice-oriented PAI work can help design, HCI and AI researchers/practitioners to navigate the field. In particular, it can serve as a reference point for reflecting on our position and standing, and for critically scrutinising the procedures we set up and questioning whether we are reinforcing power imbalances and non-equitable forms of participation.

Finally, the aim of this work is also to call our community once again to understand its political standing–to acknowledge the **deeply political nature of PAI work**. We also want to underscore the message, however, that taking a stand on these public matters does not necessarily mean to only be critical and 'against'. We would rather encourage our peers to **get inspired and celebrate the virtuous work that the community is producing, and to approach this research and practice field as a space of connection, care and collective growth**.

7.1 Reflections and Limitations

Looking back at our initial intentions and ambitions for this work, we are confronted with the fact that current PAI research, even the one committed to advancing justice, actually struggles to meet its promises. As a consequence, our initial goal to spotlight and champion existing work as a way to make methodologies and tools more accessible to the HCI community remains partially unmet. Further, we also realised that doing the very work of critically reflecting on these practices is non-trivial. PAI is an "in-situ", reflexive and holistic practice, and its complexity often cannot be captured fully by published articles only "because the implementation of AI involves navigating complex issues across multiple

organisational and political arenas and scales, including misaligned stakeholder expectations, concerns about data quality, and ethical dilemmas"([115] p.186).

Our choice of focusing on full papers only already carries the consequence of missing out on valuable work that, maybe for its ongoing nature, or the possible incompatibility with our dominant HCI publishing culture, remains documented only in short papers and late-breaking reports. Relatedly, there is a plethora of important justice-oriented work, including PAI work, being done by community-led organisations and activist groups that do not end up being published in academic venues. As researchers committed to matters of social justice, *we should build new approaches for learning from this body of non-academic work*. This could also favour the inclusion of decolonial and indigenous approaches to data work and AI discourse and development, which are starting to emerge [14, 28, 134] but still remain marginal. As a matter of fact, our work, as well as the articles in our corpus, manifest once again a W.E.I.R.D. [66] cultural hegemony even in the critical AI space of justice-oriented PAI work.

Along with a plurality of approaches to building knowledge in PAI, we should also acknowledge the diversity and specificity of challenges and implications that PAI carries, in relation to the technology being engaged with. AI, in fact, is not a univocal thing, but rather a family of technologies, each carrying particular challenges to PAI practices. As mentioned in the Introduction (1), here we use the term "AI" with quotation marks to denote its meaning in popular parlance, as suggested by the editorial guidelines of the *Critical AI journal*, and in line with the institutionalised use of the term that covers a broad range of technologies designed to solve problems traditionally assigned to human intelligence [109]. We acknowledge, however, the risks of using "AI" as a '*one-of-a-kind, ahistorical technological agent with mystical capabilities and autonomy*' [29, 122, 138] as this contributes to a 'strategic vagueness' [122] that masks actual algorithmic capabilities in favour of the marketing intentions of 'a handful of tech companies' [138]. Furthermore, within the specific case of PAI practices, different moments in the AI development pipeline come with very different challenges. If PAI is performed for developing context-specific applications of existing foundation models, as in the case of exploring opportunities for using AI in the intensive care unit [142], the challenges may be more concerned with creating a common knowledge base among stakeholders, collectively curating dedicated datasets, and more. If PAI is engaged for exploring alternative approaches to data work involved in the creation and functioning of fundational models, instead, the challenge may be more oriented towards building in annotators a sense of collective responsibility and awareness towards plural interpretations of our annotated realities, as argued by [8]. Thereafter, in this paper, we intentionally use "AI" with its generalist and popular terminology, as the ambition is to join a global critical discourse around AI that is spanning academia, industry and governmental bodies, on a theoretical level. But we encourage the community to refrain from using this generalist term when it comes to actually reporting on PAI empirical work, and to accurately describe what specific technologies were either developed or used, as this carries important implications for PAI methodologies and impact.

We acknowledge the limitations of this work, both for leaning towards Western scholarship and addressing AI in generalist terms. However, we look at this research space as one in the making, whose borders and structures are yet to be defined. So interesting questions open up, and the critical AI and HCI community, if we attempt to walk the walk of participation as justice. *What configurations can participatory alliances take in different algorithmic contexts? And how far can participatory action go in terms of affecting algorithmic development? What can be the role of governmental bodies for realising social justice ambitions through PAI work?* These are just a few of the design considerations that we can, and should, collectively engage with, if we are to build a conscious and responsible culture of participation in AI systems development.

Acknowledgments

This work is partially funded by: The Italian Ministry of University and Research, FIS 2 funding program (Project PARJAI – CUD E53C24003890001), and JEDAI: Social Justice in AI Project 20003432 ENW Diversity Initiative award.

References

[1] Amina A Abdu, Lauren M Chambers, Deirdre K Mulligan, and Abigail Z Jacobs. 2024. Algorithmic transparency and participation through the handoff lens: Lessons learned from the US Census Bureau's adoption of differential privacy. In *Proceedings of the 2024 ACM Conference on Fairness, Accountability, and Transparency*. 1150–1162.

[2] John M Abowd. 2018. The US Census Bureau adopts differential privacy. In *Proceedings of the 24th ACM SIGKDD international conference on knowledge discovery & data mining*. 2867–2867.

[3] William Agnew, A Stevie Bergman, Jennifer Chien, Mark Díaz, Seliem El-Sayed, Jaylen Pittman, Shakir Mohamed, and Kevin R McKee. 2024. The illusion of artificial inclusion. In *Proceedings of the 2024 CHI Conference on Human Factors in Computing Systems*. 1–12.

[4] William Agnew, Kevin R McKee, I Gabriel, J Kay, W Isaac, AS Bergman, S El-Sayed, and S Mohamed. 2023. Technologies of resistance to AI. *Equity and Access in Algorithms, Mechanisms, and Optimization* (2023), 1–13.

[5] Evgeni Aizenberg and Jeroen Van Den Hoven. 2020. Designing for human rights in AI. *Big Data & Society* 7, 2 (2020), 2053951720949566.

[6] Hilary Arksey and Lisa O'malley. 2005. Scoping studies: towards a methodological framework. *International journal of social research methodology* 8, 1 (2005), 19–32.

[7] Sherry R Arnstein. 1969. A ladder of citizen participation. *Journal of the American Institute of planners* 35, 4 (1969), 216–224.

[8] Anne Arzberger, Stefan Buijsman, Maria Luce Lupetti, Alessandro Bozzon, and Jie Yang. 2024. Nothing comes without its world–practical challenges of aligning llms to situated human values through RLHF. In *Proceedings of the AAAI/ACM Conference on AI, Ethics, and Society*, Vol. 7. 61–73.

[9] Ashratuz Zavin Asha, Sharifa Sultana, Helen Ai He, and Ehud Sharlin. 2024. "Shotitwo First!": Unraveling Global South Women's Challenges in Public Transport to Inform Autonomous Vehicle Design. In *Proceedings of the 2024 ACM Designing Interactive Systems Conference*. 3193–3209.

[10] Jacqui Ayling and Adriane Chapman. 2022. Putting AI ethics to work: are the tools fit for purpose? *AI and Ethics* 2, 3 (2022), 405–429.

[11] Amid Ayobi, Katarzyna Stawarz, Dmitri Katz, Paul Marshall, Taku Yamagata, Raúl Santos-Rodríguez, Peter Flach, and Aisling Ann O'Kane. 2021. Machine learning explanations as boundary objects: how AI researchers explain and non-experts perceive machine learning. (2021).

[12] Amir Baradaran. 2024. Towards a decolonial I in AI: mapping the pervasive effects of artificial intelligence on the art ecosystem. *Ai & Society* 39, 1 (2024), 7–19.

[13] Shaowen Bardzell and Jeffrey Bardzell. 2011. Towards a feminist HCI methodology: social science, feminism, and HCI. In *Proceedings of the SIGCHI conference on human factors in computing systems*. 675–684.

[14] Teanna Barrett, Chinasa T. Okolo, B. Biira, Eman Sherif, Amy Zhang, and Leilani Battle. 2025. African Data Ethics: A Discursive Framework for Black Decolonial AI. In *Proceedings of the 2025 ACM Conference on Fairness, Accountability, and Transparency (FAccT '25)*. Association for Computing Machinery, New York, NY, USA, 334–349. doi:10.1145/3715275.3732023

[15] Zack Beauchamp. 2024. Why the far right is surging all over the world. https://www.vox.com/politics/361136/far-right-authoritarianism-

germany-reactionary-spirit
[16] Eevi E Beck. 2002. P for political: Participation is not enough. *Scandinavian Journal of Information Systems* 14, 1 (2002), 1.
[17] Tom Bieling. 2019. *Design (&) activism: Perspectives on design as activism and activism as design.* Mimesis.
[18] Abeba Birhane, William Isaac, Vinodkumar Prabhakaran, Mark Diaz, Madeleine Clare Elish, Iason Gabriel, and Shakir Mohamed. 2022. Power to the people? Opportunities and challenges for participatory AI. In *Proceedings of the 2nd ACM Conference on Equity and Access in Algorithms, Mechanisms, and Optimization.* 1–8.
[19] Luigi Bobbio. 2019. Designing effective public participation. *Policy and Society* 38, 1 (2019), 41–57.
[20] Elizabeth Bondi, Lily Xu, Diana Acosta-Navas, and Jackson A Killian. 2021. Envisioning communities: a participatory approach towards AI for social good. In *Proceedings of the 2021 AAAI/ACM Conference on AI, Ethics, and Society.* 425–436.
[21] Tone Bratteteig and Ina Wagner. 2016. What is a participatory design result?. In *Proceedings of the 14th Participatory Design Conference: Full papers-Volume 1.* 141–150.
[22] Virginia Braun and Victoria Clarke. 2022. Conceptual and design thinking for thematic analysis. *Qualitative psychology* 9, 1 (2022), 3.
[23] Larasati Retno Third Aisling Brown, Venetia and Tracie Farrell. 2024. A Qualitative Study on Cultural Hegemony and the Impacts of AI. In *Proceedings of the Seventh AAAI/ACM Conference on AI, Ethics, and Society (AIES2024).* 226–238.
[24] Jude Browne, Stephen Cave, Eleanor Drage, and Kerry McInerney. 2023. *Feminist AI: Critical perspectives on algorithms, data, and intelligent machines.* Oxford University Press.
[25] Alessandra Buccella. 2023. "AI for all" is a matter of social justice. *AI and Ethics* 3, 4 (2023), 1143–1152.
[26] Carrie J Cai, Samantha Winter, David Steiner, Lauren Wilcox, and Michael Terry. 2021. Onboarding materials as cross-functional boundary objects for developing ai assistants. In *Extended Abstracts of the 2021 CHI Conference on Human Factors in Computing Systems.* 1–7.
[27] Luciano Cavalcante Siebert, Maria Luce Lupetti, Evgeni Aizenberg, Niek Beckers, Arkady Zgonnikov, Herman Veluwenkamp, David Abbink, Elisa Giaccardi, Geert-Jan Houben, Catholijn M Jonker, et al. 2023. Meaningful human control: actionable properties for AI system development. *AI and Ethics* 3, 1 (2023), 241–255.
[28] Marianela Ciolfi Felice, Ivana Feldfeber, Carolina Glasserman Apicella, Yasmín Belén Quiroga, Julián Ansaldo, Luciano Lapenna, Santiago Bezchinsky, Raul Barriga Rubio, and Mailén García. 2025. Doing the Feminist Work in AI: Reflections from an AI Project in Latin America. In *Proceedings of the 2025 CHI Conference on Human Factors in Computing Systems (CHI '25).* Association for Computing Machinery, New York, NY, USA, Article 998, 18 pages. doi:10.1145/3706598.3713681
[29] Marianela Ciolfi Felice, Ivana Feldfeber, Carolina Glasserman Apicella, Yasmín Belén Quiroga, Julián Ansaldo, Luciano Lapenna, Santiago Bezchinsky, Raul Barriga Rubio, and Mailén García. 2025. Doing the Feminist Work in AI: Reflections from an AI Project in Latin America. In *Proceedings of the 2025 CHI Conference on Human Factors in Computing Systems.* 1–18.
[30] Patricia Hill Collins. 2022. *Black feminist thought: Knowledge, consciousness, and the politics of empowerment.* routledge.
[31] Patricia Hill Collins and Sirma Bilge. 2020. *Intersectionality.* John Wiley & Sons.
[32] Patricia Hill Collins, Elaini Cristina Gonzaga da Silva, Emek Ergun, Inger Furseth, Kanisha D Bond, and Jone Martínez-Palacios. 2021. Intersectionality as critical social theory: Intersectionality as critical social theory, Patricia Hill Collins, Duke University Press, 2019. *Contemporary Political Theory* 20, 3 (2021), 690.
[33] Eric Corbett, Remi Denton, and Sheena Erete. 2023. Power and public participation in AI. In *Proceedings of the 3rd ACM Conference on Equity and Access in Algorithms, Mechanisms, and Optimization.* 1–13.
[34] Sasha Costanza-Chock. 2018. Design justice: Towards an intersectional feminist framework for design theory and practice. In *Proceedings of the DRS International Conference 2018, 25-28 June, Limerick, Ireland.* 529–540.
[35] Sasha Costanza-Chock. 2020. *Design justice: Community-led practices to build the worlds we need.* The MIT Press.
[36] Fernando Delgado, Stephen Yang, Michael Madaio, and Qian Yang. 2023. The participatory turn in ai design: Theoretical foundations and the current state of practice. In *Proceedings of the 3rd ACM Conference on Equity and Access in Algorithms, Mechanisms, and Optimization.* 1–23.
[37] Tawanna R Dillahunt, Alex Jiahong Lu, and Joanna Velazquez. 2023. Eliciting alternative economic futures with working-class detroiters: Centering afrofuturism in speculative design. In *Proceedings of the 2023 ACM Designing Interactive Systems Conference.* 957–977.
[38] Carl DiSalvo. 2015. *Adversarial design.* Mit Press.
[39] C DiSalvo, A Clement, and V Pipek. 2013. Communities: Participatory Design for, with and by Communities,[w:] Routledge International Handbook of Participatory Design, red. J. Simonsen, T. Robertson.
[40] DJN. 2018. Design Justice Network Principles. https://designjustice.org/read-the-principles
[41] DJN. 2025. Design Justice Network. https://designjustice.org/about-us
[42] Sibel Duz. 2025. Gaza as a testing ground: Israel's AI warfare. https://www.setav.org/en/gaza-as-a-testing-ground-israels-ai-warfare
[43] Elizabeth Edenberg and Alexandra Wood. 2023. Disambiguating algorithmic bias: From neutrality to justice. In *Proceedings of the 2023 AAAI/ACM Conference on AI, Ethics, and Society.* 691–704.
[44] Journal Editors. 2025. For Authors. https://dukeupress.edu/critical-ai#ForAuthors
[45] Upol Ehsan, Ranjit Singh, Jacob Metcalf, and Mark Riedl. 2022. The algorithmic imprint. In *Proceedings of the 2022 ACM Conference on Fairness, Accountability, and Transparency.* 1305–1317.
[46] Salma Elsayed-Ali, Elizabeth Bonsignore, and Joel Chan. 2023. Exploring challenges to inclusion in participatory design from the perspectives of global north practitioners. *Proceedings of the ACM on Human-Computer Interaction* 7, CSCW1 (2023), 1–25.
[47] Motahhare Eslami, Sarah Fox, Hong Shen, Bobbie Fan, Yu-Ru Lin, Rosta Farzan, and Beth Schwanke. 2025. From Margins to the Table: Charting the Potential for Public Participatory Governance of Algorithmic Decision Making. In *Proceedings of the 2025 ACM Conference on Fairness, Accountability, and Transparency.* 2657–2670.
[48] Gregory Falco. 2019. Participatory AI: Reducing AI bias and developing socially responsible AI in smart cities. In *2019 IEEE International Conference on Computational Science and Engineering (CSE) and IEEE International Conference on Embedded and Ubiquitous Computing (EUC).* IEEE, 154–158.
[49] Sina Fazelpour and Zachary C Lipton. 2020. Algorithmic fairness from a non-ideal perspective. In *Proceedings of the AAAI/ACM Conference on AI, Ethics, and Society.* 57–63.
[50] Michael Feffer, Michael Skirpan, Zachary Lipton, and Hoda Heidari. 2023. From preference elicitation to participatory ML: A critical survey & guidelines for future research. In *Proceedings of the 2023 AAAI/ACM Conference on AI, Ethics, and Society.* 38–48.
[51] Laura E Forlano and Megan K Halpern. 2023. Speculative histories, just futures: From counterfactual artifacts to counterfactual actions. *ACM Transactions on Computer-Human Interaction* 30, 2 (2023), 1–37.
[52] Nick J Fox. 2011. Boundary objects, social meanings and the success of new technologies. *Sociology* 45, 1 (2011), 70–85.
[53] Sarah Fox, Mariam Asad, Katherine Lo, Jill P Dimond, Lynn S Dombrowski, and Shaowen Bardzell. 2016. Exploring social justice, design, and HCI. In *Proceedings of the 2016 CHI Conference Extended Abstracts on Human Factors in Computing Systems.* 3293–3300.
[54] Nancy Fraser. 2006. *Mapping the Feminist Imagination: From Redistribution to Recognition to Representation.* VS Verlag für Sozialwissenschaften, Wiesbaden, 37–51. doi:10.1007/978-3-531-90382-8_2
[55] Andreas Fuster, Paul Goldsmith-Pinkham, Tarun Ramadorai, and Ansgar Walther. 2022. Predictably unequal? The effects of machine learning on credit markets. *The Journal of Finance* 77, 1 (2022), 5–47.
[56] Ian G Johnson and Clara Crivellaro. 2021. Opening research commissioning to civic participation: creating a community panel to review the social impact of HCI research proposals. In *Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems.* 1–17.
[57] Iason Gabriel. 2022. Toward a theory of justice for artificial intelligence. *Daedalus* 151, 2 (2022), 218–231.
[58] Ben Gansky and Sean McDonald. 2022. CounterFAccTual: How FAccT undermines its organizing principles. In *Proceedings of the 2022 ACM Conference on Fairness, Accountability, and Transparency.* 1982–1992.
[59] Aakash Gautam. 2024. Reconfiguring participatory design to resist ai realism. In *Proceedings of the Participatory Design Conference 2024: Exploratory Papers and Workshops-Volume 2.* 31–36.
[60] Mathieu Edouard Samborska Veronika Giattino, Charlie and Max Roser. 2023. Data Page: Number of large-scale AI systems released per year. https://archive.ourworldindata.org/20250813-134018/grapher/number-of-large-scale-ai-systems-released-per-year.html
[61] Pauline Gourlet, Donato Ricci, and Maxime Crépel. 2024. Reclaiming artificial intelligence accounts: A plea for a participatory turn in artificial intelligence inquiries. *Big Data & Society* 11, 2 (2024), 20539517241248093.
[62] Brett A Halperin, Gary Hsieh, Erin McElroy, James Pierce, and Daniela K Rosner. 2023. Probing a community-based conversational storytelling agent to document digital stories of housing insecurity. In *Proceedings of the 2023 CHI Conference on Human Factors in Computing Systems.* 1–18.
[63] MD Romael Haque, Devansh Saxena, Katy Weathington, Joseph Chudzik, and Shion Guha. 2024. Are we asking the right questions?: Designing for community stakeholders' interactions with ai in policing. In *Proceedings of the 2024 CHI Conference on Human Factors in Computing Systems.* 1–20.
[64] Christina Harrington, Sheena Erete, and Anne Marie Piper. 2019. Deconstructing community-based collaborative design: Towards more equitable participatory design engagements. *Proceedings of the ACM on human-computer interaction* 3,

CSCW (2019), 1–25.
[65] Steve Harrison, Phoebe Sengers, and Deborah Tatar. 2011. Making epistemological trouble: Third-paradigm HCI as successor science. *Interacting with computers* 23, 5 (2011), 385–392.
[66] Joseph Henrich, Steven J Heine, and Ara Norenzayan. 2010. The weirdest people in the world? *Behavioral and brain sciences* 33, 2-3 (2010), 61–83.
[67] Hans-Martin Heyn and Eric Knauss. 2022. Structural causal models as boundary objects in AI system development. In *Proceedings of the 1st International Conference on AI Engineering: Software Engineering for AI*. 43–45.
[68] AI Hleg. 2019. Ethics guidelines for trustworthy AI. *B-1049 Brussels* (2019).
[69] Isto Huvila, Theresa Dirndorfer Anderson, Eva Hourihan Jansen, Pam McKenzie, and Adam Worrall. 2017. Boundary objects in information science. *Journal of the Association for Information Science and Technology* 68, 8 (2017), 1807–1822.
[70] IxDF. 2023. Participatory Design. https://www.interaction-design.org/literature/topics/participatory-design
[71] Petra Jääskeläinen, André Holzapfel, and Emelie Eriksson. 2024. AI Art for Self-Interest or Common Good? Uncovering Value Tensions in Artists' Imaginaries of AI Technologies. In *Proceedings of the 2024 ACM Designing Interactive Systems Conference*. 2897–2910.
[72] Tobi Jegede, Marissa Kumar Gerchick, Amreeta S Mathai, and Aaron Horowitz. 2023. Challenge accepted? A critique of the 2021 National Institute of Justice recidivism forecasting challenge. In *Proceedings of the 3rd ACM Conference on Equity and Access in Algorithms, Mechanisms, and Optimization*. 1–17.
[73] Jean-Marie John-Mathews. 2022. Some critical and ethical perspectives on the empirical turn of AI interpretability. *Technological Forecasting and Social Change* 174 (2022), 121209.
[74] Jean-Marie John-Mathews, Dominique Cardon, and Christine Balagué. 2022. From reality to world. A critical perspective on AI fairness. *Journal of Business Ethics* 178, 4 (2022), 945–959.
[75] Guy Julier. 2013. From design culture to design activism. *Design and Culture* 5, 2 (2013), 215–236.
[76] Anna Kawakami, Amanda Coston, Haiyi Zhu, Hoda Heidari, and Kenneth Holstein. 2024. The situate AI Guidebook: Co-designing a toolkit to support multi-stakeholder, early-stage deliberations around public sector AI proposals. In *Proceedings of the 2024 CHI Conference on Human Factors in Computing Systems*. 1–22.
[77] Jackie Kay, Atoosa Kasirzadeh, and Shakir Mohamed. 2024. Epistemic injustice in generative AI. In *Proceedings of the AAAI/ACM Conference on AI, Ethics, and Society*, Vol. 7. 684–697.
[78] Finn Kensing and Jeanette Blomberg. 1998. Participatory design: Issues and concerns. *Computer supported cooperative work (CSCW)* 7, 3 (1998), 167–185.
[79] Hannah Rose Kirk, Bertie Vidgen, Paul Röttger, and Scott A Hale. 2023. Personalisation within bounds: A risk taxonomy and policy framework for the alignment of large language models with personalised feedback. *arXiv preprint arXiv:2303.05453* (2023).
[80] Lauren Klein and Catherine D'Ignazio. 2024. Data feminism for AI. In *Proceedings of the 2024 ACM Conference on Fairness, Accountability, and Transparency*. 100–112.
[81] Xander Koo, Lucy Scott, and Amy Bruckman. 2025. Metrics and Macchiatos: Challenges for Service-Industry Wdustry Workers and the Need for Worker-Driven ICTs. *ACM Transactions on Computer-Human Interaction* (2025).
[82] Tanja Kubes. 2025. A Critique of Human-Centred AI: A Plea for a Feminist AI Framework (FAIF). *AI & SOCIETY* (2025), 1–12.
[83] Sebastian Linxen, Christian Sturm, Florian Brühlmann, Vincent Cassau, Klaus Opwis, and Katharina Reinecke. 2021. How weird is CHI?. In *Proceedings of the 2021 chi conference on human factors in computing systems*. 1–14.
[84] Thomas Lodato and Carl DiSalvo. 2018. Institutional constraints: the forms and limits of participatory design in the public realm. In *Proceedings of the 15th Participatory Design Conference: Full Papers-Volume 1*. 1–12.
[85] Maria Luce Lupetti. 2022. Understanding critical design tensions with progression through historical practices. *Diid—Disegno Industriale Industrial Design* 77 (2022), 12.
[86] Maria Luce Lupetti and Dave Murray-Rust. 2024. (un) making ai magic: a design taxonomy. In *Proceedings of the 2024 CHI Conference on Human Factors in Computing Systems*. 1–21.
[87] Maria Isabel Magaña and Katie Shilton. 2025. Frameworks, Methods and Shared Tasks: Connecting Participatory AI to Trustworthy AI Through a Systematic Review of Global Projects. In *Proceedings of the 2025 ACM Conference on Fairness, Accountability, and Transparency*. 2166–2179.
[88] Thomas Markussen. 2013. The disruptive aesthetics of design activism: enacting design between art and politics. *Design Issues* 29, 1 (2013), 38–50.
[89] Afonso Matos. 2022. *Who Can Afford to be Critical?: An Inquiry Into what We Can't Do Alone, as Designers, and Into what We Might be Able to Do Together, as People*. Set Margins'.
[90] Dan McQuillan. 2022. *Resisting AI: An anti-fascist approach to artificial intelligence*. Policy Press.
[91] David Miller. 2021. Justice. *Stanford Encyclopedia of Philosophy* (2021).
[92] Michael J Muller and Sarah Kuhn. 1993. Participatory design. *Commun. ACM* 36, 6 (1993), 24–28.
[93] Jimin Mun, Liwei Jiang, Jenny Liang, Inyoung Chung, Nicole DeCario, Yejin Choi, Tadayoshi Kohno, and Maarten Sap. 2025. *Particip-AI: A Democratic Surveying Framework for Anticipating Future AI Use Cases, Harms and Benefits*. AAAI Press, 997–1010.
[94] N/A. 2024. With Global Crises Increasing Levels of Poverty, Unprecedented Investments Will Be Needed to Reverse Trend, Speakers Tell Second Committee. https://press.un.org/en/2024/gaef3607.doc.htm
[95] N/A. 2025. Facts about the nature crisis. https://www.unep.org/facts-about-nature-crisis
[96] N/A. 2025. Tech giants see emissions surge 150 percent in 3 years amid AI boom: UN. https://www.aljazeera.com/news/2025/6/6/tech-giants-see-emissions-surge-150-percent-in-3-years-amid-ai-boom-un
[97] Angella Ndaka, Philippe JC Lassou, Konan Anderson Seny Kan, and Samuel Fosso-Wamba. 2024. Toward response-able AI: A decolonial perspective to AI-enabled accounting systems in Africa. *Critical Perspectives on Accounting* 99 (2024), 102736.
[98] Julia Nee, Genevieve Macfarlane Smith, Alicia Sheares, and Ishita Rustagi. 2021. Advancing social justice through linguistic justice: Strategies for building equity fluent NLP technology. In *Proceedings of the 1st ACM Conference on Equity and Access in Algorithms, Mechanisms, and Optimization*. 1–9.
[99] Ma Michelle Nicoletti, Leonardo and Dina Bass. 2025. AI is draining water from areas that need it most. https://www.bloomberg.com/graphics/2025-ai-impacts-data-centers-water-data/
[100] Noel Pearse. 2019. An illustration of deductive analysis in qualitative research. In *18th European conference on research methodology for business and management studies*. 264.
[101] Adrian Petterson, Keith Cheng, and Priyank Chandra. 2023. Playing with power tools: Design toolkits and the framing of equity. In *Proceedings of the 2023 CHI Conference on Human Factors in Computing Systems*. 1–24.
[102] Jennifer Pierre, Roderic Crooks, Morgan Currie, Britt Paris, and Irene Pasquetto. 2021. Getting Ourselves Together: Data-centered participatory design research & epistemic burden. In *Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems*. 1–11.
[103] Organizers Of Queerinai, Anaelia Ovalle, Arjun Subramonian, Ashwin Singh, Claas Voelcker, Danica J Sutherland, Davide Locatelli, Eva Breznik, Filip Klubicka, Hang Yuan, et al. 2023. Queer in AI: A case study in community-led participatory AI. In *Proceedings of the 2023 ACM Conference on Fairness, Accountability, and Transparency*. 1882–1895.
[104] Ben Quinn and Dan Milmo. 2024. How the far right is weaponising AI-generated content in Europe. https://www.theguardian.com/technology/2024/nov/26/far-right-weaponising-ai-generated-content-europe
[105] John Rawls. 2022. *Political Liberalism*. Columbia University Press.
[106] Ingrid Robeyns and Morten Fibieger Byskov. 2025. The Capability Approach. In *The Stanford Encyclopedia of Philosophy* (Summer 2025 ed.), Edward N. Zalta and Uri Nodelman (Eds.). Metaphysics Research Lab, Stanford University.
[107] María Méndez Rocasolano. 2022. Human rights, big data and artificial intelligence: Elements of a complex algorithm. In *Security and Defence: Ethical and Legal Challenges in the Face of Current Conflicts*. Springer, 93–102.
[108] Elizabeth B-N Sanders and Pieter Jan Stappers. 2008. Co-creation and the new landscapes of design. *Co-design* 4, 1 (2008), 5–18.
[109] Filippo Santoni de Sio, Txai Almeida, and Jeroen Van Den Hoven. 2024. The future of work: freedom, justice and capital in the age of artificial intelligence. *Critical Review of International Social and Political Philosophy* 27, 5 (2024), 659–683.
[110] Laura Schelenz. 2025. Black Feminism (BF) and Artificial Intelligence (AI): The Possibilities and Limitations of Contesting Discriminatory AI from a Critical Social Theory Perspective. *Frontiers in Sociology* 10 (2025), 1602947.
[111] Andrew D Selbst, Danah Boyd, Sorelle A Friedler, Suresh Venkatasubramanian, and Janet Vertesi. 2019. Fairness and abstraction in sociotechnical systems. In *Proceedings of the conference on fairness, accountability, and transparency*. 59–68.
[112] Laleh Seyyed-Kalantari, Haoran Zhang, Matthew BA McDermott, Irene Y Chen, and Marzyeh Ghassemi. 2021. Underdiagnosis bias of artificial intelligence algorithms applied to chest radiographs in under-served patient populations. *Nature medicine* 27, 12 (2021), 2176–2182.
[113] Mona Sloane. 2024. Controversies, contradiction, and "participation" in AI. *Big Data & Society* 11, 1 (2024), 20539517241235862.
[114] Mona Sloane, Emanuel Moss, Olaitan Awomolo, and Laura Forlano. 2022. Participation is not a design fix for machine learning. In *Proceedings of the 2nd ACM Conference on Equity and Access in Algorithms, Mechanisms, and Optimization*. 1–6.
[115] Rachel Charlotte Smith, Liesbeth Huybrechts, Jesper Simonsen, and Daria Loi. 2025. Contemporary Participatory Design: Research Agendas for Societal Crisis. In *Proceedings of the sixth decennial Aarhus conference: Computing X Crisis*. 182–201.
[116] Jaemarie Solyst, Shixian Xie, Ellia Yang, Angela EB Stewart, Motahhare Eslami, Jessica Hammer, and Amy Ogan. 2023. "I would like to design": Black girls

analyzing and ideating fair and accountable AI. In *Proceedings of the 2023 CHI Conference on Human Factors in Computing Systems*. 1–14.

[117] Clay Spinuzzi. 2005. The methodology of participatory design. *Technical communication* 52, 2 (2005), 163–174.

[118] Bernd Carsten Stahl. 2021. Ethical issues of AI. In *Artificial Intelligence for a better future: An ecosystem perspective on the ethics of AI and emerging digital technologies*. Springer, 35–53.

[119] Susan Leigh Star and James R Griesemer. 1989. Institutional ecology,translations' and boundary objects: Amateurs and professionals in Berkeley's Museum of Vertebrate Zoology, 1907-39. *Social studies of science* 19, 3 (1989), 387–420.

[120] Paul Stares. 2025. Conflicts to Watch in 2025. https://www.cfr.org/report/conflicts-watch-2025

[121] Marc Steen. 2013. Co-design as a process of joint inquiry and imagination. *Design issues* 29, 2 (2013), 16–28.

[122] Lucy Suchman. 2023. The uncontroversial 'thingness' of AI. *Big Data & Society* 10, 2 (2023), 20539517231206794.

[123] Sharifa Sultana, Hafsah Mahzabin Chowdhury, Zinnat Sultana, and Nervo Verdezoto. 2025. 'Socheton': A Culturally Appropriate AI Tool to Support Reproductive Well-being. In *Proceedings of the 2025 ACM Designing Interactive Systems Conference*. 3098–3116.

[124] Harini Suresh, Emily Tseng, Meg Young, Mary Gray, Emma Pierson, and Karen Levy. 2024. Participation in the age of foundation models. In *Proceedings of the 2024 ACM Conference on Fairness, Accountability, and Transparency*. 1609–1621.

[125] Mariarosaria Taddeo and Luciano Floridi. 2018. How AI can be a force for good. *Science* 361, 6404 (2018), 751–752.

[126] Tiera Tanksley, Angela DR Smith, Saloni Sharma, and Earl W Huff Jr. 2025. " Ethics is not neutral": Understanding Ethical and Responsible AI Design from the Lenses of Black Youth. In *Proceedings of the 2025 CHI Conference on Human Factors in Computing Systems*. 1–20.

[127] Jordan Taylor and Amy Bruckman. 2024. Mitigating epistemic injustice: The online construction of a bisexual culture. *ACM Transactions on Computer-Human Interaction* 31, 4 (2024), 1–34.

[128] Anja Thieme, Maryann Hanratty, Maria Lyons, Jorge Palacios, Rita Faia Marques, Cecily Morrison, and Gavin Doherty. 2023. Designing human-centered AI for mental health: Developing clinically relevant applications for online CBT treatment. *ACM Transactions on Computer-Human Interaction* 30, 2 (2023), 1–50.

[129] Johannes Thumfart. 2024. The democratic offset: Contestation, deliberation, and participation regarding military applications of AI. *AI and Ethics* 4, 2 (2024), 511–526.

[130] Sophie Toupin. 2024. Shaping feminist artificial intelligence. *New Media & Society* 26, 1 (2024), 580–595.

[131] Loreben Tuquero. 2024. Did artificial intelligence shape the 2024 US election? https://www.aljazeera.com/news/2024/12/25/did-artificial-intelligence-shape-the-2024-us-election

[132] Barbara-Chiara Ubaldi and Ricardo Zapata. 2024. GOVERNING WITH ARTIFICIAL INTELLIGENCE: ARE GOVERNMENTS READY? https://www.oecd.org/content/dam/oecd/en/publications/reports/2024/06/governing-with-artificial-intelligence_f0e316f5/26324bc2-en.pdf

[133] Bao-Thi Van Cong, Séamus A Power, and Thomas A Morton. 2025. Solidarity with whom? Minority perspectives on allyship in Danish queer spaces. *British Journal of Social Psychology* 64, 1 (2025), e12780.

[134] Kush R. Varshney. 2025. *Decolonial AI Alignment: Openness, Viśeundefineda-Dharma, and Including Excluded Knowledges*. AAAI Press, 1467–1481.

[135] Danding Wang, Qian Yang, Ashraf Abdul, and Brian Y Lim. 2019. Designing theory-driven user-centric explainable AI. In *Proceedings of the 2019 CHI conference on human factors in computing systems*. 1–15.

[136] Salla Westerstrand. 2024. Reconstructing AI ethics principles: Rawlsian ethics of artificial intelligence. *Science and Engineering Ethics* 30, 5 (2024), 1–21.

[137] Cedric Deslandes Whitney, Teresa Naval, Elizabeth Quepons, Simrandeep Singh, Steven R Rick, and Lilly Irani. 2021. HCI tactics for politics from below: Meeting the challenges of smart cities. In *Proceedings of the 2021 CHI conference on human factors in computing systems*. 1–15.

[138] Meredith Whittaker. 2021. The steep cost of capture. *Interactions* 28, 6 (2021), 50–55.

[139] Meredith Whittaker, Meryl Alper, Cynthia L Bennett, Sara Hendren, Liz Kaziunas, Mara Mills, Meredith Ringel Morris, Joy Rankin, Emily Rogers, Marcel Salas, et al. 2019. Disability, bias, and AI. *AI Now Institute* 8 (2019).

[140] Christine T Wolf. 2020. Democratizing AI? experience and accessibility in the age of artificial intelligence. *XRDS: Crossroads, The ACM Magazine for Students* 26, 4 (2020), 12–15.

[141] Wenjun Yang and Eyhab Al-Masri. 2018. Democratizing Differential Privacy: A Participatory AI Framework for Public Decision-Making. *arXiv preprint arXiv:2504.21297v2* (2018).

[142] Nur Yildirim, Susanna Zlotnikov, Deniz Sayar, Jeremy M Kahn, Leigh A Bukowski, Sher Shah Amin, Kathryn A Riman, Billie S Davis, John S Minturn, Andrew J King, et al. 2024. Sketching ai concepts with capabilities and examples: ai innovation in the intensive care unit. In *Proceedings of the 2024 CHI conference on human factors in computing systems*. 1–18.

[143] Meg Young, Upol Ehsan, Ranjit Singh, Emnet Tafesse, Michele Gilman, Christina Harrington, and Jacob Metcalf. 2024. Participation versus scale: Tensions in the practical demands on participatory AI. *First Monday* (2024).

[144] Cristina Zaga and Maria Luce Lupetti. 2022. *Diversity equity and inclusion in embodied AI: reflecting on and re-imagining our future with embodied AI*. 4TU. Federation.

[145] Douglas Zytko, Pamela J. Wisniewski, Shion Guha, Eric PS Baumer, and Min Kyung Lee. 2022. Participatory design of AI systems: opportunities and challenges across diverse users, relationships, and application domains. In *CHI Conference on Human Factors in Computing Systems Extended Abstracts*. 1–4.